Large Language Models, Price Discovery, and the Limits to Adoption

Jiawei Wang*†

April 2026

Abstract

Large language models (LLMs) are increasingly used as informational intermediaries in financial markets. This paper asks whether they improve market efficiency, whether these gains are monotonic as adoption spreads, and whether they narrow the performance gap between institutional and retail investors. I develop a model in which LLM adoption improves the informativeness of order flow but also increases correlated trading errors, generating a hump-shaped relation between adoption and market efficiency. Using exogenous ChatGPT outages as quasi-natural experiments, I show that variance ratios temporarily rise during outages, implying slower price discovery. I then use outage-based variation to construct firm-level ChatGPT exposure and show that more exposed firms experience larger post-launch efficiency gains, but that these gains are nonmonotonic and peak about ten months after launch. Finally, retail trading profitability improves relative to institutional investors, suggesting that LLMs reduce traditional informational asymmetries.

Keywords: large language models, market efficiency, price discovery, informational intermediation
JEL Classification: G10, D40, O30

*Warwick Business School, University of Warwick. Email:Jiawei.Wang.2@warwick.ac.uk

†This paper was previously circulated under the title Large Language Models and Stock Price Discovery. It has greatly benefited from discussions with Hengjie Ai, Constantinos Antoniou, Patrick Coen, Gaofeng Han, Po-Hsuan Hsu, Tony Hu, Ruggero Jappelli, Yan Ji, Olga Klein, Roman Kozhan, John Kuong, Noah Lyman, Kebin Ma, Henry Mak, Philippe Mueller, Ruslan Sverchkov, Giorgio Valente, Ganesh Viswanath-Natraj, Junxuan Wang, Zhengge Zhou, and seminar participants at HKMA, 2025 FMA Doctoral Student Consortium, Warwick Business School Brown Bag, and the Gillmore Center Fintech PhD Workshop 2025.

1 Introduction

Large language models (LLMs) are becoming an increasingly important layer of informational intermediation in financial markets. Much like earlier advances in trading and information technologies (e.g., Hendershott et al. (2011); Brogaard et al. (2014)), LLMs can affect how efficiently public information is incorporated into prices. They serve as information processing tools for analysts and investors (Blankespoor et al., 2024; Bertomeu et al., 2025), facilitating trading decisions (Cheng et al., 2025), and even demonstrating return predictive power (Lopez-Lira and Tang, 2023). Survey evidence indicates that nearly half of professional investors already rely on generative AI tools, primarily to interpret financial data and simplify complex disclosures (Ebert, 2023). Recent industry reports likewise document the rapid adoption of LLMs by both institutional and retail investors in financial decision making (Alves, 2025). By lowering the cost of processing earnings announcements, 10-K filings, financial news, and other rich textual sources, LLMs have the potential to accelerate price discovery. At the same time, widespread reliance on the similar black-box models may generate correlated errors and synchronized trading, introducing new sources of fragility (International Monetary Fund, 2024). This raises a natural set of questions: Do LLMs improve market efficiency? If so, does efficiency increase monotonically with adoption, or is it *too much of a good thing?* And does broader access to LLM-based information processing narrow the performance gap between institutional and retail investors?

To organize these forces, I develop a parsimonious model in the spirit of Kyle (1985). In the model, LLM adoption changes the informational environment of the market in two distinct ways. First, it increases the informativeness of non-insider order flow by helping traders extract signals from public information. In this sense, LLMs act as a new form of informational intermediation: they do not generate new fundamentals, but enhance investors' ability to process existing information and trade on it. Second, when many traders rely on similar models, prompts, or data pipelines, LLM adoption increases the correlation of trading errors. Order flow therefore becomes not only more informative, but also more synchronized. Unlike Kyle (1985), I allow the market maker to face inventory costs. This feature matters because correlated order imbalances affect prices not only through information revelation but also through inventory pressure. As a result, the same technology that improves information processing can also generate price distortions

when adoption becomes sufficiently widespread.

The model yields three testable predictions. First, LLM adoption improves market efficiency by increasing the informativeness of order flow. Second, this effect is non-monotonic: at low levels of adoption, informational gains dominate, but as adoption becomes widespread, correlated errors and inventory pressure offset these gains, generating a hump-shaped relation between adoption and market efficiency. Third, by lowering information-processing costs, LLMs reduce the informational disadvantage of retail investors and narrow the performance gap between retail and institutional traders.

I test these predictions using high-frequency data and several complementary empirical designs. My baseline measure of firm-level market efficiency is the variance ratio (Bessembinder, 2003; Rösch et al., 2017), where higher values indicate greater departures from a random walk and thus slower price discovery. The first set of tests exploits plausibly exogenous ChatGPT outages as shocks to LLM availability[1]. This approach follows prior work in market microstructure that exploits service disruptions to identify the effects of information frictions (Shive, 2012; Cheng et al., 2025). If LLMs help investors process public information, then temporary outages should slow price discovery. In total, I identify 39 outage events during U.S. trading hours over the period from November 2022 to December 2024. These incidents are largely driven by provider-side technical problems, such as infrastructure failures, software bugs, and capacity constraints, rather than by shocks in financial markets. For each outage, I identify the affected five-minute intervals and compare them with two matched control sets: the same clock-time intervals in the five trading days before and after the outage, and unaffected five-minute intervals within the same trading day. For each treatment and control interval, I compute the variance ratio and compare the resulting estimates. The difference between the two captures the causal effect of the outage. To separate informational efficiency from contemporaneous trading frictions, I further orthogonalize the variance ratio with respect to five minute trading volume and the quoted spread. The residual from this procedure serves as an alternative proxy for market efficiency that is purged of liquidity driven variation.

Using these measures, I show that variance ratios at the five-minute horizon rise during outages relative to both control groups, providing causal evidence that LLMs im-

[1]ChatGPT is the focus of the analysis because it was the first widely adopted high-capability LLM, which provides the longest time series for empirical study, and it remains the dominant chatbot product: https://gs.statcounter.com/ai-chatbot-market-share.

prove the speed at which new information is incorporated into prices. Applying the Within–Between decomposition of Mundlak (1978), I further find that these efficiency gains are most pronounced for liquid, high-turnover firms, for stocks with greater return volatility, and for firms with higher institutional ownership and more complex disclosures. Higher volatility points to larger efficiency benefits precisely where price discovery is typically more difficult, while tighter spreads and higher turnover indicate that LLM-generated insights are incorporated most rapidly in markets with low frictions and high trading intensity. Greater institutional ownership is associated with larger improvements, consistent with institutional investors being better positioned to process and act on LLM-generated insights (Blankespoor et al., 2024), and firms with more complex disclosures display greater gains, suggesting that AI tools add the most value when human processing costs are high.

A natural concern is whether investors would respond to outages that occur at such a high frequency. The evidence suggests they do: trading volume drops significantly during ChatGPT outages, and the decline is concentrated in non-retail trading. By contrast, retail volume is relatively stable at the 5-minute horizon, consistent with retail investors being less exposed to high-frequency execution and monitoring. Sophisticated participants, such as algorithmic traders and other non-retail investors who may rely on the ChatGPT API or interface, reduce activity when that input becomes unavailable. The interpretation does not require that all such investors rely on ChatGPT: it is sufficient that a marginal subset of traders pause, delay, or curtail trading during outages. Moreover, even if some participants substitute toward alternative or local LLMs, substitution is unlikely to be uniform; heterogeneity in tools and model outputs can increase dispersion in the signals traders extract from text. Importantly, the stability of retail trading should not be read as evidence that retail investors do not use LLMs. Rather, it suggests that retail trading is less immediately responsive to short-horizon service disruptions. Even outside high-frequency strategies, both retail and institutional investors trade in any given five-minute interval for routine reasons (e.g., liquidity needs or portfolio maintenance). In that sense, outages remove a widely used "information filter" precisely when some subset of market participants is making trading decisions.

Another concern is that five-minute intervals will rarely align exactly with the timestamps of major scheduled disclosures (e.g., 10-K or 10-Q releases), so it may be unclear how outages at this horizon could affect informational efficiency. The key is that LLM usage

is not confined to the instant a quarterly report is released. First, markets process a near-continuous stream of text-based information: news updates, macro releases, firm announcements, and social-media content. And investors use ChatGPT to summarize and interpret these inputs on demand. Second, even when information is released at a discrete time, attention and processing are dispersed: investors consult LLMs when they actually contemplate trades, not necessarily at the release minute. This staggered decision making implies variation over time in the set of traders relying on ChatGPT, even if any single investor trades infrequently. For these reasons, outages can plausibly impair informational efficiency across many five-minute intervals, not only at the exact moments of major announcements, which is exactly the type of short-horizon deviation the variance-ratio approach is designed to capture.

I next turn to the model's second prediction: the efficiency gains from LLM adoption need not be monotonic over time. To study this prediction, I use the outage-based evidence to construct a firm-level measure of ChatGPT trading exposure, defined by the extent to which a stock's trading volume falls during outage intervals relative to matched control periods. Treating the public launch of ChatGPT as an exogenous structural break, I estimate a generalized difference-in-differences specification that compares changes in market efficiency across firms with different levels of ChatGPT exposure before and after the launch. The results show that market efficiency improves significantly after the launch, with larger gains for more exposed firms.

I then trace the dynamics of this effect using an event-study specification with three-month bins relative to the launch date. The estimates reveal a pronounced hump-shaped pattern. Efficiency improves in the initial post-launch period, increases further as the public LLM ecosystem expands, and reaches its strongest effect roughly 6 to 12 months after launch, before partially reversing. The reversal is concentrated among firms with greater ChatGPT exposure, consistent with the view that LLM-assisted information processing is most beneficial at intermediate stages of adoption. A statistical test based on a quadratic post-event profile confirms this non-monotonicity and places the turning point at about 10.5 months after launch. Taken together, these results support the core prediction of the model: LLM adoption initially enhances market quality, but the marginal gains decline and eventually partially unwind as adoption becomes broader and trading behavior becomes more correlated.

Finally, I test the model's prediction on performance gap across investor types. I iden-

tify retail trades using the price-improvement algorithm of Boehmer et al. (2021), and institutional trades using the large-trade classifier of Lee and Radhakrishna (2000). Following Barber et al. (2024), I construct standardized order-imbalance measures for each group and form daily long-short portfolios that buy stocks in the highest imbalance quintile and short those in the lowest. These portfolios track the profitability of retail- and institution-led trades before and after the LLM launch and across firms with different levels of ChatGPT exposure. I find that retail trading profitability improves in the LLM era, particularly among high-exposure firms, while institutional benchmark strategies lose part of their advantage. A spread portfolio that is long the retail strategy and short the institutional strategy earns higher returns after the launch of ChatGPT, indicating that the relative performance of retail investors has improved. Taken together, these results suggest that broader access to powerful information-processing tools can narrow the traditional performance gap between institutions and individuals.

This paper makes at least three main contributions. First, it brings LLMs into market microstructure as a new form of informational intermediation and provides causal, high-frequency evidence that LLM availability improves the speed of price discovery. While prior work has used comparable shocks to study long-horizon informativeness (e.g., Cheng et al. (2025)), this paper adopts a market structure perspective, using variance ratios to capture how quickly information is incorporated into prices rather than whether it eventually is. Second, this is the first paper to provide empirical evidence of a non-linear, hump-shaped evolution of market efficiency gains following the adoption of LLM, and it develops a theoretical framework that rationalizes this pattern by modeling the trade-off between the informational benefits of adoption and the systemic risks of correlated trading. Third, it contributes to the debate on distributional effect of AI (e.g., Guo et al. (2022); Chang et al. (2023)) by providing direct evidence that LLMs act as a democratizing force, narrowing the realized performance gap between retail and institutional traders. More broadly, the findings suggest that the benefits of LLM adoption depend not only on the extent of adoption, but also on model diversity, signal quality, and the resilience of markets to synchronized order flow.

The remainder of the paper proceeds as follows. Section 2 reviews the related literature. Section 3 presents the theoretical framework. Section 4 describes the data and measures. Section 5 presents the empirical evidence. Section 6 discusses the alternative measures. Section 7 concludes.

2 Related Literature

This paper relates to a broader literature on information and market efficiency (Fama, 1970; Grossman and Stiglitz, 1980; Kyle, 1985). Prior work shows that sophisticated investors and new trading technologies can improve the efficiency with which information is incorporated into prices (Hendershott et al., 2011; Boehmer and Kelley, 2009; Shive, 2012; Brogaard et al., 2014). In this context, LLMs can be viewed as a technology that lowers the cost of extracting signals from publicly available information and thereby shifts some market participants toward more informed trading (Chang et al., 2023). My findings support this information improving role of LLMs by showing that ChatGPT adoption enhances market efficiency. At the same time, the effect is not monotonic. This feature connects the paper to work emphasizing that a greater presence of sophisticated or algorithmic traders need not always improve market efficiency. For example, when traders rely on similar unanchored strategies, uncertainty about how many others are using the same signal can generate crowding and price overshooting (Stein, 2009). Price informativeness can be non-monotonic in investor size due to information pass through versus learning channels (Kacperczyk et al., 2025). AI-powered trading may also sustain inefficient collusive equilibria (Dou et al., 2025). This paper complements these studies by documenting and rationalizing a hump-shaped relation between LLM adoption and market efficiency.

An emerging body of research shows that LLMs are powerful tools for processing the large volume of textual information that underlies financial markets. Studies demonstrate their ability to extract and analyze information from corporate disclosures such as 10-K filings and earnings calls (Shaffer and Wang, 2024; Serafeim, 2024; Beckmann et al., 2024; Jha et al., 2024), public information sources such as financial news and social media (Lopez-Lira and Tang, 2023; Huang et al., 2024; Chen et al., 2024, 2025), and more specialized texts including analyst reports, employee reviews, government statements, startup descriptions, and narrative corporate filings (Lv, 2024; Hansen and Kazinnik, 2024; Hu et al., 2025; Li et al., 2026; Cao et al., 2026). GPT-generated summaries of complex financial disclosures can better predict returns by more effectively capturing decision-relevant content (Kim et al., 2023). This paper builds on this literature by moving beyond document-level text processing to examine broader market consequences. I link the information-processing capabilities of LLMs to measurable market outcomes, docu-

menting improvements in high-frequency market efficiency, attenuation of post-earnings announcement drift, and a narrowing of the performance gap between retail and institutional investors.

A related strand of literature exploits external shocks to LLM access to identify their market consequences. For example, Cheng et al. (2025) document significant declines in trading volume during major ChatGPT outages and report long-run gains in price informativeness associated with ChatGPT-assisted trading. Bertomeu et al. (2025) use Italy's temporary ChatGPT ban to show that constrained access leads to less accurate analyst forecasts and market-wide declines in informational efficiency. This paper complements these studies along two dimensions. First, whereas prior work often emphasizes long-run price informativeness, I adopt a market-structure perspective and focus on high-frequency variance ratios that capture the speed of price discovery. Second, I characterize the dynamic evolution of these efficiency gains, showing that they are non-linear and vary systematically with the level of LLM adoption.

This paper also contributes to a growing literature on the distributional and systemic consequences of AI in financial markets. Chang et al. (2023) show that ChatGPT's release narrowed information gaps between retail investors and institutional short sellers, suggesting a democratization of sophisticated analysis. Yet adoption remains uneven: hedge fund reliance on AI rises most among large, active, and high-performing funds, with greater adoption linked to better performance (Sheng et al., 2024); sophisticated users integrate generative AI sooner than novices (Blankespoor et al., 2024); and AI-based investment consultants can improve individual outcomes unevenly across investor types (Guo et al., 2022). I extend this literature by moving beyond information gap proxies to track realized trading performance directly through long-short strategies that mimic retail and institutional order imbalances.

More broadly, the paper speaks to concerns that widespread reliance on AI may induce correlated beliefs, crowd trading, and fragile market dynamics (International Monetary Fund, 2024; Stein, 2009). These risks are magnified when models generate correlated errors, such as hallucinations (Ji et al., 2023; Kang and Liu, 2023; Lo and Ross, 2024; Zhang et al., 2025; Xu et al., 2025). I contribute to this debate by providing both empirical evidence of a hump-shaped efficiency dynamic and a theoretical framework in which the information-processing benefits of LLMs are eventually offset by the systemic costs of correlated trading.

3 Model

This section develops a simple model of LLM adoption in financial markets. The model embeds LLM adoption in a Kyle (1985) framework in which greater adoption increases the informativeness of non-insider order flow, but also raises the correlation of trading errors across market participants. The resulting tradeoff generates a hump-shaped relation between adoption and price efficiency. Unlike Kyle (1985), I allow the market maker to face inventory costs, which enter the pricing rule and capture a realistic feature of financial markets. In this setting, correlated trading errors and inventory frictions jointly generate price inefficiency.

3.1 Environment

Consider a Kyle (1985) market with one strategic informed trader, a mass of non-insider traders, and a competitive market maker. The terminal value of the risky asset is

$$v \sim \mathcal{N}(0, \sigma_v^2).$$

Aggregate order flow is

$$y = x + u,$$

where x denotes the informed trader's order and u denotes aggregate non-insider order flow.

Non-insiders observe public information relevant to fundamentals, but without LLM assistance they cannot process it precisely enough to trade systematically on that information. For example, a trader may lack the financial expertise or analytical tools needed to extract a sufficiently precise trading signal from public disclosures, earnings calls, or news. In this sense, their orders are dominated by liquidity demand, and the environment reduces to the noise trader benchmark in Kyle (1985). LLM adoption does not expand the information set. Instead, it improves the precision with which existing public information is processed and translated into trading decisions.

Economically, this creates two types of informed trading. The first is the strategic informed trader in Kyle (1985), who internalizes price impact and chooses trading aggres-

siveness optimally. The second is an LLM-assisted trader whose order becomes correlated with fundamentals because the trader can process public information more effectively, but who is not fully strategic and therefore does not adjust trading aggressiveness in response to market depth in the same way as the informed trader. At the same time, because adopters may rely on similar models, prompts, or data environments, their trading errors may become correlated. Thus, LLM adoption does not create new information. Rather, it improves the processing of existing public information while also creating the possibility of correlated mistakes across adopters. In this sense, the model can be interpreted as a Kyle (1985) market with one strategic informed trader and an additional non-strategic source of informative order flow.

I capture this channel in reduced form by assuming that aggregate non-insider order flow is

$$u = \varepsilon + A\phi v + A\eta,$$

where $A \in [0, 1]$ denotes the adoption intensity of the non-insider sector. The term $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ represents residual liquidity trading. The term $A\phi v$ captures the idea that LLM-assisted traders extract information from public signals in a way that is positively related to fundamentals. The parameter $\phi > 0$ measures the extent to which LLM adoption improves the precision of public signal processing. The term $A\eta$, where $\eta \sim \mathcal{N}(0, \sigma_\eta^2)$, captures a common model error shared across adopters. The random variables v, ε, and η are mutually independent.

After observing aggregate order flow y, the market maker sets the price p. Unlike in Kyle (1985), the market maker faces inventory costs. For tractability, these costs are assumed to be quadratic:

$$C(y) = \delta y^2, \qquad \delta > 0,$$

3.2 Linear Equilibrium

Following Kyle (1985), I focus on a linear equilibrium. The market maker observes aggregate order flow y and sets a price $p(y)$. Given y, the competitive market maker satisfies the zero-profit condition

$$\mathbb{E}\big[(p(y) - v)y - C(y) \mid y\big] = 0.$$

I assume that the market maker does not separately observe the AI-related component of non-insider demand, and instead conditions only on aggregate order flow. Under this assumption, AI affects prices indirectly through its impact on the informativeness and error structure of order flow. This assumption is natural for three reasons. First, it preserves the standard structure of Kyle (1985), in which prices respond to information revealed through trading. Second, although AI may be widely available, the market maker need not observe the realized AI-driven trading impulse separately from aggregate demand. Third, this formulation keeps the model focused on the central question of how LLM adoption alters the informational content of order flow. If the market maker observed the AI component directly, the model would instead become one in which AI-processed information is incorporated into prices outside the trading process.

The informed trader observes v and chooses x to maximize expected trading profits,

$$\max_{x} \ \mathbb{E}\big[(v-p)x \mid v\big].$$

One may also wonder why the informed trader does not condition her strategy on the AI-generated component embedded in non-insider order flow. In principle, she would have an incentive to do so. Because she already observes the fundamental value, such a signal would not improve her estimate of fundamentals directly, but it could help her decompose aggregate order flow and trade against the error component of LLM-assisted non-insider demand. I abstract from this channel for two reasons. First, it is not necessary for the main mechanism of the model, which is to capture how LLM adoption alters the correlation between non-insider order flow and fundamentals, as well as the possibility of correlated mistakes. Second, as emphasized by Goldstein et al. (2025), access to such coarse information can generate equilibrium effects that may even reduce the well-informed trader's payoff. I abstract from that margin here in order to keep the model focused on how LLM adoption changes the informativeness and error structure of non-insider order flow.

Equilibrium definition. A linear equilibrium consists of an insider demand function $x = \beta(A)v$ and a pricing rule $p = \lambda(A)y$ such that the insider's order is optimal given the pricing rule, and the pricing rule satisfies the market maker's zero-profit condition given the insider's strategy.

Define

$$B(A) \equiv \beta(A) + A\phi,$$

as the total sensitivity of aggregate order flow to fundamentals. It combines the strategic informed trader's contribution, $\beta(A)$, with the AI-induced fundamental component of non-insider demand, $A\phi$. Under a linear insider strategy, aggregate order flow can then be written as

$$y = B(A)v + \varepsilon + A\eta.$$

Proposition 1. *A linear equilibrium is characterized by*

$$x = \beta(A)v, \qquad p = \lambda(A)y,$$

where

$$\beta(A) = \frac{1 - \lambda(A)A\phi}{2\lambda(A)}, \qquad B(A) \equiv \beta(A) + A\phi, \qquad \lambda(A) = \delta + \frac{B(A)\sigma_v^2}{B(A)^2\sigma_v^2 + \sigma_\varepsilon^2 + A^2\sigma_\eta^2}.$$

The expression for $\lambda(A)$ has a natural interpretation. The first term, δ, reflects compensation for inventory costs, while the second term captures the informational component of price impact. When aggregate order flow becomes more informative about fundamentals, the market maker raises price impact in response to stronger adverse-selection concerns. As a result, equilibrium price impact is jointly determined by market-making frictions and the informativeness of trading. A higher $\lambda(A)$ implies a less aggressive trading response by the strategic informed trader, since she internalizes that larger orders move prices more strongly against her.

The equilibrium can be summarized through the scalar $B(A)$, given that

$$\lambda(A) = \frac{1}{2B(A) - A\phi},$$

it is convenient to define

$$G(B, A) \equiv \frac{1}{2B - A\phi} - \delta - \frac{B\sigma_v^2}{B^2\sigma_v^2 + \sigma_\varepsilon^2 + A^2\sigma_\eta^2}.$$

An equilibrium branch is a function $B(A)$ satisfying

$$G(B(A), A) = 0.$$

Proposition 2. *Suppose that along the relevant equilibrium branch,*

$$\frac{2}{(2B - A\phi)^2} > \frac{\sigma_v^2\big(B^2\sigma_v^2 - (\sigma_\varepsilon^2 + A^2\sigma_\eta^2)\big)}{\big(B^2\sigma_v^2 + \sigma_\varepsilon^2 + A^2\sigma_\eta^2\big)^2}.$$

Then $B'(A) > 0$. The same condition is sufficient for local uniqueness of the equilibrium branch.

The proposition shows that a higher adoption rate raises the fundamental component of aggregate order flow, as long as the noise component is not too large relative to the signal component. This does not yet imply that market efficiency improves monotonically, because the same increase in adoption also enlarges the common error component $A\eta$.

3.3 Market Efficiency

Define market inefficiency as the mean squared pricing error

$$M(A) \equiv \mathbb{E}\big[(v - p)^2\big].$$

Let

$$S(A) \equiv B(A)^2\sigma_v^2, \qquad N(A) \equiv \sigma_\varepsilon^2 + A^2\sigma_\eta^2.$$

Then

$$\mathrm{Var}(y) = S(A) + N(A).$$

Here $S(A)$ is the informational component of order flow, while $N(A)$ is the non-fundamental component.

Proposition 3. *Suppose $B'(A) > 0$ on $[0, 1]$. Then*

$$M(A) = \sigma_v^2\frac{N(A)}{S(A) + N(A)} + \delta^2\big(S(A) + N(A)\big).$$

If

$$\delta^2 < \sigma_v^2\frac{\sigma_\varepsilon^2}{(S(0) + \sigma_\varepsilon^2)^2} \quad \textit{and} \quad \delta^2\big(S'(1) + 2\sigma_\eta^2\big) > \sigma_v^2\frac{(\sigma_\varepsilon^2 + \sigma_\eta^2)S'(1) - 2\sigma_\eta^2 S(1)}{(S(1) + \sigma_\varepsilon^2 + \sigma_\eta^2)^2},$$

then there exists at least one $A^* \in (0,1)$ *such that*

$$M'(A^*) = 0.$$

Hence market efficiency is hump-shaped in the adoption rate.

This proposition formalizes the central tradeoff in the model. Market inefficiency $M(A)$ has two components. The first is a learning component, given by the posterior variance of fundamentals conditional on order flow. It falls when order flow becomes more informative about fundamentals, but rises when a larger share of order flow reflects non-fundamental trading. The second is an inventory component, which increases with total order-flow volatility.

Common model error worsens efficiency through both channels. First, it makes aggregate order flow less informative, because comovement in trading increasingly reflects correlated misinterpretation rather than genuine information about fundamentals. Second, it raises the volatility of order flow and therefore the inventory risk borne by the market maker. Since the market maker cannot perfectly distinguish informed trading from correlated non-fundamental demand, this additional volatility generates stronger price pressure and larger pricing errors.

The proposition therefore explains why the effect of adoption on efficiency need not be monotone. At low levels of adoption, the main effect of LLM use is to improve the informational content of non-insider order flow, so efficiency rises. At high levels of adoption, however, correlated errors become more important, and inventory costs magnify the resulting distortions. Market efficiency can therefore be hump-shaped in the adoption rate: some adoption improves price discovery, but too much adoption can reduce it.

Because the equilibrium loading $B(A)$ is endogenous, closed-form comparative statics with respect to primitive parameters are not especially transparent. I therefore study the efficiency profile numerically. Figure 1 plots market efficiency, measured by $1/M(A)$, as a function of the adoption rate A. Three patterns emerge.

First, lower inventory costs increase market efficiency at all adoption levels. When δ is smaller, the market maker absorbs order imbalances with less price distortion, so the entire efficiency curve shifts upward. In the limit $\delta \to 0$, the model converges to the Kyle (1985), in which the strategic interaction between informed traders and the market

maker endogenously offsets the distortions generated by correlated LLM signals. As a result, efficiency can increase monotonically with adoption.

Second, lower common model error improves market efficiency and attenuates the decline in efficiency at high adoption levels. A reduction in σ_η reduces the extent to which LLM adopters make correlated mistakes, thereby lowering the non-fundamental component of aggregate order flow. Numerically, this flattens the right side of the hump and shifts the efficient range of adoption to the right.

Third, greater informational loading improves market efficiency throughout. A higher value of ϕ strengthens the component of order flow aligned with fundamentals, which enhances price discovery across the full range of adoption rates.

3.4 Distributional Effect

A transparent measure of the distributional effect of adoption is the informed trader's expected profit,

$$\Pi_I(A) \equiv \mathbb{E}[(v-p)x].$$

In this setting, trading gains are redistributed across market participants. Because the market maker earns zero expected profit net of inventory compensation, the trading game is effectively zero-sum up to the inventory cost wedge. As a result, lower insider profits reflect a weaker informational advantage and a narrower gap between informed and uninformed traders.

Proposition 4. *For a given adoption rate A, insider profit is given by*

$$\Pi_I(A) = \frac{\sigma_v^2}{4\lambda(A)}\big(1-\lambda(A)A\phi\big)^2.$$

The expression for $\Pi_I(A)$ clarifies why adoption is likely to reduce insider rents. As A rises, the component $A\phi v$ makes non-insider order flow more closely aligned with fundamentals. This reduces the exclusivity of the insider's information and compresses the insider's trading advantage. In other words, adoption shifts part of the information incorporated into prices away from the insider and into aggregate order flow. Although adoption also affects market depth through $\lambda(A)$, and this channel may in principle

attenuate or amplify insider rents, the erosion of the insider's informational edge provides a natural force pushing $\Pi_I(A)$ downward.

Figure 2 illustrates this mechanism numerically. Across all three panels, insider profit declines with the adoption rate, indicating that greater adoption tends to reduce the insider's informational rents. The left panel shows that this pattern holds for different values of the inventory cost δ. Lower inventory frictions raise the level of insider profit, but they do not overturn the downward effect of adoption. The middle panel varies the common model error σ_η. Insider profit again declines with adoption, although the slope becomes flatter when common model error is larger, reflecting the fact that correlated mistakes partly offset the loss of the insider's advantage. The right panel varies the informational loading ϕ. A larger ϕ leads insider profit to fall more steeply with adoption, consistent with the idea that more informative non-insider trading more rapidly erodes the insider's informational edge.

Taken together, the numerical results suggest that LLM adoption reduces the insider's ability to extract rents from private information. In this sense, adoption not only affects aggregate market efficiency, but also has an important distributional consequence: it narrows the performance gap between informed traders and the uninformed side. At the same time, the comparative statics reveal that the forces shaping market efficiency and the forces shaping the distribution of trading profits do not always move in the same direction. In particular, lower inventory frictions improve market efficiency by reducing price distortions, but they also tend to raise insider profit by allowing the informed trader to trade more aggressively on private information. By contrast, lower common model error and higher informational loading both improve the quality of non-insider order flow and reduce insider rents. More broadly, improvements in model quality and financial tailoring can both strengthen price discovery and weaken the distributional advantage of informed traders by making AI-assisted trading more informative about fundamentals (Chen et al., 2025).

3.5 Empirical Predictions

The model yields three empirical predictions that guide the subsequent analysis[2].

[2]It is worth noting that model parameters such as $\phi, \sigma_\eta, \delta$ can shape the equilibrium relations, as shown in Figures 1 and 2. In principle, these comparative statics generate further empirical predictions.

Prediction 1: LLM adoption improves market efficiency. At low levels of adoption, LLMs increase the informativeness of non-insider order flow, making prices more reflective of fundamentals. Accordingly, greater LLM adoption improves market efficiency.

Prediction 2: The effect on market efficiency is nonmonotonic. As adoption becomes widespread, correlated model errors become more important. This raises the non-fundamental component of order flow and, in the presence of inventory costs, amplifies price distortions, implying a hump-shaped relation between LLM adoption and market efficiency.

Prediction 3: LLM adoption improves relative performance of retail investors. By lowering information-processing costs, LLMs reduce the informational advantage of sophisticated traders, narrowing the performance gap between institutional and retail investors.

4 Data

Sample Construction. The sample comprises U.S. common stocks (CRSP share codes 10 and 11) listed on the NYSE, AMEX, and NASDAQ. The analysis spans from January 2021 through December 2024, focusing on regular trading hours (09:30–16:00 EST). This sample period is selected to bracket the public launch of ChatGPT (November 30, 2022), providing balanced pre- and post-event windows. Tick-by-tick transactions are obtained from TAQ, while daily stock prices are drawn from CRSP. Observations with stock prices below $5 are excluded to avoid distortions from microcap stocks.

Pricing Efficiency Measures. Stock pricing efficiency is measured using the variance ratio (e.g., Bessembinder (2003); Rösch et al. (2017)), which evaluates the extent to which individual stock prices follow a random walk. For stock i on day d, the variance ratio is defined as

Since they are not the central focus of the paper, I do not pursue them empirically here.

$$VR_{id} = \left| \frac{\text{Var}(Ret_{id}^{1\text{min}})}{4 \times \text{Var}(Ret_{id}^{15\text{sec}})} - 1 \right|$$

where $\text{Var}(Ret_{id}^{1\text{min}})$ and $\text{Var}(Ret_{id}^{15\text{sec}})$ denote the variances of one-minute and fifteen-second log returns, respectively, within day d[3]. A higher value of VR_{id} indicates a larger deviation from the random walk benchmark and thus lower pricing efficiency.

The variance ratio is particularly informative in high-frequency settings because it captures short-horizon return reversals relative to longer-horizon return variation. Under an efficient market with frictionless trading, returns should approximate a random walk, yielding a variance ratio close to zero; deviations therefore reflect transitory frictions or delays in information incorporation. The choice of fifteen-second versus one-minute returns follows prior work, as it balances capturing rich short-horizon price information with avoiding excessive sensitivity to microstructure noise.[4]

LLM Outages Events. ChatGPT serves as the focal point of this analysis because it dominates the current market[5] and was the first widely adopted, high capability LLM, providing the longest available time series for empirical test. Following recent work (e.g., Cheng et al. (2025)), outage events are used as plausibly exogenous shocks to identify the causal effect of ChatGPT on financial markets. A comprehensive set of outage incidents from launch through the end of 2024 is compiled from OpenAI's public archive of service disruptions.[6] For each event, the start and end times are recorded. The sample is filtered to include only events where: (i) the title references *GPT*, (ii) the component status indicates an *outage* and (iii) the outage at least partially overlaps with U.S. trading hours. This filtering process yields 39 distinct outage events (detailed in Table 1), with an average duration of 84 minutes. These incidents are mostly driven by technical problems on the provider side, such as infrastructure failures, software bugs, and capacity constraints, and affect all users of the service contemporaneously, rather than being triggered by shocks in financial markets. Moreover, the timing of outages does not systematically line up with

[3] In the outage test, VR definition is adapted to the 5-minute observational window: the numerator and denominator represent the variances of one-minute and fifteen-second log returns, respectively, computed within each five-minute interval.

[4] Similar results obtain when using alternative horizons such as 5 seconds, 5 minutes, 15 minutes, or a PCA-based combination of variance ratios.

[5] See https://gs.statcounter.com/ai-chatbot-market-share.

[6] See https://status.openai.com/history.

major macroeconomic announcements or firm-specific news releases, supporting the view that they are orthogonal to underlying information flows.

For each of the 39 outage events, I collect all five-minute interval observations that fall within the affected trading window, together with the corresponding intervals from the five trading days before and after the outage. I also retain all five-minute intervals on the outage days themselves. This sampling design yields two natural comparison groups. The first, denoted *Control Group I*, consists of the same clock-time five-minute intervals in the ± 5 trading days when no outage occurred. The second, denoted *Control Group II*, consists of unaffected intervals within the same trading day[7].

The variance ratio is then computed for each treatment and control interval. To isolate the informational component of efficiency from trading frictions, I orthogonalize the variance ratio with respect to five-minute trading volume and the quoted spread. The resulting measure, denoted $VRorth$, serves as an alternative proxy for market efficiency.

Control Variables. A range of firm- and market-level characteristics are included as controls to account for potential determinants of pricing efficiency: firm-level variables include price, return, size, and trading volume from CRSP; institutional ownership from LSEG 13F filings; effective spreads and trade-based return volatility from TAQ; and the number of firm-specific information events in the preceding 90 days from Capital IQ Key Developments; textual complexity of corporate filings is measured using the Flesch–Kincaid Grade Level, obtained from the SEC Analytics Suite via WRDS. At the market level, the VIX index is included to capture prevailing aggregate uncertainty.[8]

Summary Statistics. Table 2 reports descriptive statistics for the main variables. All variables are winsorized at the 1% tails. Trading-related variables such as VR, $VRorth$, effective spread, volatility, price, and turnover exhibit pronounced right skewness, with means substantially exceeding medians and large maxima reflecting occasional bursts of illiquidity or extreme trading activity. In contrast, institutional ownership shows a wide

[7]For example, during the outage on March 15, 2023 from 14:14 to 14:20, the treatment intervals are 14:10–14:15 and 14:15–14:20. Control Group I comprises the same intervals in the surrounding ± 5 trading days (March 8–10, 13–14, and 16–17, 21–22), excluding March 20 where the same intervals coincided with another outage. Control Group II consists of all other five-minute intervals on March 15, namely 09:30–14:10 and 14:20–16:00.

[8]Collected from https://fred.stlouisfed.org/series/VIXCLS.

but relatively balanced distribution, ranging from near zero to almost complete ownership. Textual complexity varies more moderately, with average Flesch–Kincaid scores around 20, corresponding to a graduate-level reading difficulty.

5 Empirical Results

5.1 Model Prediction 1: Market Efficiency

For both control group comparisons, I estimate the market efficiency test using the following specification:

$$VR_{itd} = \beta\, Outage_t + \mathbf{X}_{i,d-1}\delta + \alpha_i + \gamma_t + \varepsilon_{it}, \tag{1}$$

where i, t, and d index firm, five-minute interval, and date, respectively. The coefficient of interest, β, captures the difference in efficiency between outage and control periods. A positive β implies that variance ratios are higher during outages, indicating lower efficiency and suggesting that ChatGPT improves the speed of price discovery under normal conditions. The vector $\mathbf{X}_{i,d-1}$ contains firm-level control variables lagged by one day to mitigate simultaneity concerns and to serve as instruments for contemporaneous values. Following prior work (e.g., Rösch et al. (2017); Boehmer and Kelley (2009)), the lagged dependent variable is also included to account for the persistence of efficiency measures and to ensure that autocorrelation does not bias the estimates. Firm fixed effects (α_i) and 5-minute interval fixed effects (γ_t) further control for unobserved heterogeneity.

A natural concern is that five-minute intervals will rarely coincide exactly with major scheduled disclosures such as 10-K or 10-Q filings, raising the question of how outages at this horizon can affect efficiency. The key is that LLM is not only used at the moment of quarterly reports. Financial markets are subject to a near-continuous flow of information such as news releases, macro announcements, firm-specific updates, and social-media signals, and ChatGPT is routinely used as an on-demand filter and summarization tool for these rich textual sources. Even for periodic disclosures, investors do not all read and process the information at the exact release time: they consult LLMs when they

are actually considering trades (for example, when rebalancing portfolios or after receiving income). This staggered usage pattern implies at least some variation in the set of traders relying on ChatGPT across five-minute intervals, even if individual investors do not adjust positions that frequently. As a result, outages are expected to degrade informational efficiency broadly across five-minute intervals, not only at the precise moments of major announcements, which is precisely what the variance-ratio approach is designed to capture.

Table 3 reports the regression results. All specifications include control variables and fixed effects. The coefficients on *Outage* are significantly positive at the 1% level across all columns, indicating that stock pricing efficiency decreases during outages. The economic magnitude, corresponding to roughly 1% ($\frac{0.294}{0.33\times100}$) of a standard deviation in variance ratios, is also meaningful. This estimate should be viewed as a conservative lower bound on the true economic contribution of LLMs. It is derived from a temporary disruption to only one platform (ChatGPT), while the rest of the generative AI ecosystem including competitor models and proprietary institutional tools remained fully operational. Moreover, as this 1% effect applies to a single five-minute interval, the cumulative impact on price discovery over a typical 84-minute outage (the mean) is substantial. The fact that a partial, short-lived shock to a single tool yields a statistically detectable degradation in market efficiency underscores the significant role these tools already play in the price discovery process.

Market efficiency displays persistence, as indicated by the positive coefficient on the lagged efficiency term. Greater liquidity, reflected in a lower effective spread, is associated with higher efficiency, while higher institutional ownership also predicts more efficient pricing, consistent with prior evidence (e.g., Boehmer and Kelley (2009)). By contrast, the coefficient on *complexity* is significantly positive, implying that firms with harder-to-read filings exhibit lower efficiency. This pattern suggests that LLMs such as ChatGPT may be particularly valuable in helping investors process complex disclosures.

Taken together, these results indicate that ChatGPT improves market efficiency, which supports the model prediction 1. However, the improvement is unlikely to be uniform across firms. It is therefore important to identify which types of firms benefit most from these efficiency gains.

To explore this heterogeneity, the average change in variance ratios between outage and

control periods is computed for each event. To account for variation across events, firms are ranked within each outage by the magnitude of their variance ratio increase, and then assigned to deciles from 1 (lowest) to 10 (highest). A higher rank denotes a larger increase in variance ratio (i.e., a greater loss of efficiency) during the outage. This procedure is implemented using both control groups, yielding two measures of efficiency change: dVR_I^{rank} (outage vs. Control Group I) and dVR_{II}^{rank} (outage vs. Control Group II). To explain the cross-sectional variation in dVR^{rank}, I relate these measures to one-day lagged firm characteristics using the Within–Between decomposition of Mundlak (1978), which separates cross-sectional from within-firm effects.

$$dVR_{i,o}^{\text{rank}} \;=\; \left(\mathbf{X}_{i,d(o)} - \bar{\mathbf{X}}_i\right)\boldsymbol{\beta}_{\text{within}} \;+\; \bar{\mathbf{X}}_i\,\boldsymbol{\beta}_{\text{between}} \;+\; \mu_o \;+\; \varepsilon_{i,o}, \tag{2}$$

where $\mathbf{X}_{i,d(o)}$ is the vector of firm i's characteristics on the day d corresponding to outage event o, and $\bar{\mathbf{X}}_i$ is the firm-level mean of these characteristics across all events. The term μ_o denotes outage-time fixed effects. In this decomposition, the coefficients $\boldsymbol{\beta}_{\text{between}}$ capture which persistent firm characteristics explain why some firms are systematically more affected than others, while $\boldsymbol{\beta}_{\text{within}}$ measure how deviations of a firm's characteristics from its own average influence its sensitivity to outages. The main focus is on the between-firm coefficients, $\boldsymbol{\beta}_{\text{between}}$. A pure fixed effects model (firm FE + event FE) would absorb all between-firm variation, eliminating the cross-sectional heterogeneity of interest, whereas pooled OLS conflates within- and between-firm variation and can overstate significance. The Mundlak approach balances these by jointly estimating within- and between-firm effects while properly accounting for event-time heterogeneity.

Table 4 reports the results. Several persistent firm characteristics are systematically associated with larger efficiency gains during outage episodes. First, firms with higher return volatility exhibit larger gains, implying that efficiency benefits are particularly important in riskier firms where price discovery is typically more difficult. Second, firms with tighter bid–ask spreads and higher turnover are also more significantly affected by the outages. While high liquidity stocks tend to be less volatile, these characteristics proxy for a different economic mechanism: the speed of information impounding. This finding suggests that LLM-generated insights are more rapidly and effectively integrated into prices in markets with low frictions and high trading intensity. Thus, the results are not contradictory. Instead, they imply that the benefits of LLMs are largest for firms

that are both informationally complex (high volatility) and trade in an environment where new information is rapidly acted upon (high liquidity/turnover). Thirdly, greater institutional ownership is also associated with larger improvements, which may reflect that institutional investors are better positioned to process and act on ChatGPT outputs. Finally, firms with more complex disclosures display greater gains, consistent with the idea that AI tools add more value when human processing costs are high.

5.2 Model Prediction 2: Dynamic Effect of AI Adoption

5.2.1 AI Exposure

I exploit the decline in trading volume during ChatGPT outages to measure firm level exposure to AI assisted trading, following Cheng et al. (2025). Table 5 shows that trading volume falls significantly during ChatGPT outages. The decline is concentrated in non-retail trading.[9] This pattern suggests that retail investors are less responsive to high-frequency disruptions at the five-minute horizon, whereas non-retail participants, including trading bots and other sophisticated investors that may rely on the ChatGPT API or interface, reduce trading activity sharply when access is interrupted. Importantly, this does not imply that retail investors do not use LLMs; rather, their trading appears less sensitive to short-horizon service disruptions.

Figure 3 illustrates the pronounced decline in aggregate trading activity during outage intervals relative to matched control periods. I construct a firm-level measure of ChatGPT trading exposure based on this decline. For each outage event, I compute the change in trading volume between the outage interval and its matched control interval, and multiply the difference by -1, so that higher values indicate greater reliance on ChatGPT. To account for heterogeneity across outage events, I rank firms into deciles within each event based on the magnitude of their volume decline, and then average these ranks across events to obtain the final exposure measure. This measure is higher for firms with greater institutional ownership, analyst coverage, return volatility, and financial leverage (Cheng et al., 2025).

[9]Retail trading is identified using the algorithm of Boehmer et al. (2021), while non-retail trading volume is defined as total trading volume minus retail trading volume.

5.2.2 Launch of ChatGPT

The launch of ChatGPT is treated as an exogenous shock. Because adoption was staggered across users and applications, a relatively long window is considered: November 30, 2022 to December 31, 2024 defines the post-shock period, and January 1, 2021 to November 30, 2022 serves as the pre-shock period, ensuring a balanced sample. Using the ChatGPT exposure measure constructed in the previous section, a generalized difference-in-differences framework is estimated:

$$VR_{id} = \beta \left(Post_d \times Exposure_i \right) + Post_d + Exposure_i + X_{i,d-1}\delta + \varepsilon_{id}. \tag{3}$$

The coefficient β captures the differential change in efficiency for firms with higher ChatGPT exposure in the post-shock period, relative to less-exposed firms. The results, reported in Table 6, confirm the hypothesis. While Column (1) shows an average, market-wide efficiency improvement after the launch, Column (2) introduces the interaction. The coefficient on $Post \times Exposure$ is negative and statistically significant, indicating that firms with greater exposure experience disproportionately larger efficiency improvements following the launch.

5.2.3 AI Adoption and Market Efficiency

In the model, A denotes the adoption rate, but it can also be interpreted more loosely as a reduced-form index of calendar time, with higher values corresponding to broader adoption of LLM technology. This interpretation is broadly consistent with observed trends in practice, as illustrated in Figure 4[10]. Under this interpretation, adoption rate is assumed to be growing over time after the launch of ChatGPT.

To trace the evolution of this effect, an event-study specification is estimated, with event time defined in 3-month relative to the launch date:

[10]Because the true adoption rate is difficult to observe directly, the Google Trends index provides a useful proxy for the rising interest in and adoption of LLM technology.

$$VR_{id} = \sum_{k=-6}^{6} \beta_k \, \mathbf{1}\{k \neq -1,\, event\_time = k\} \times Exposure_i + X'_{i,d-1}\delta + \alpha_i + \mu_{mo(d)} + \varepsilon_{id}. \quad (4)$$

Event time is grouped into 3-month bins ranging from $k = -6$ to $k = +6$, where the $k = -6$ bin pools all months ≤ -18 (18+ months before launch) and the $k = +6$ bin pools all months $\geq +18$ (18+ months after). The month bin immediately preceding the launch ($k = -1$, covering months -3 to -1) serves as the reference period. Each event-time indicator is interacted with the firm's ChatGPT exposure measure. Firm fixed effects and month fixed effects ($\alpha_i + \mu_{mo(d)}$) absorb unobserved heterogeneity across firms and common time shocks. In this framework, the coefficients β_k trace the differential evolution of efficiency for more- versus less-exposed firms relative to the pre-launch benchmark. A negative estimate of β_k implies that firms with greater exposure experience larger improvements in efficiency at event time k.

Figure 5 plots the dynamic treatment effects (the coefficients β_k from Equation 4). Before the launch (blue area), the coefficients are generally indistinguishable from zero. In the first three months after launch (green area, corresponding to the initial phase of the public LLM ecosystem, when GPT-3.5 is the main deployed model), efficiency improves significantly relative to the reference period. This early post-launch effect is statistically significant but smaller than in the subsequent phase, consistent with early output inaccuracies and staggered adoption across market participants. After roughly three months (orange area), as the broader LLM ecosystem deepens with wider adoption and the introduction of more capable models such as GPT-4 and competing systems, market efficiency rises more sharply. The gains peak roughly 6–12 months post-launch and then partially reverse. The subsequent decline in efficiency gains suggests diminishing marginal returns to LLM adoption over time.

Although several pre-launch coefficients are statistically significant, they do not indicate a systematic pre-trend in the relation between ChatGPT exposure and market efficiency. The relevant identifying question is whether more exposed firms were already experiencing differential changes in efficiency prior to the public release of ChatGPT. The event-study evidence provides little support for this interpretation. In the pre-launch period, the coefficients are modest, fluctuate around zero, and show no persistent directional

pattern across bins. In contrast, the post-launch coefficients are consistently negative and economically larger, implying that market efficiency improves disproportionately more for firms with greater ChatGPT exposure after the shock. Taken together, the evidence points to a discrete post-launch shift in the exposure–efficiency relation rather than a continuation of pre-existing differential trends.

To formally verify the U-shaped dynamic in the market inefficiency, I estimate the following regression:

$$VR_{id} = \sum_{k \leq -1}^{6} \beta_k \, \mathbf{1}\{event\_time = k\} \times Exposure_i + \gamma_1 k_{post} \times Exposure_i + \gamma_2 k_{post}^2 \times Exposure_i + X'_{i,d-1}\delta + \alpha_i + \mu_{mo(d)} + \varepsilon_{id}. \quad (5)$$

where $k_{\text{post}} = \max(k, 0)$ is truncated below at 0. Table 7 shows that the post-event path is U-shaped in event time: the coefficient on the linear term is $\gamma_1 = -0.315$ ($t = -26.142$) and the coefficient on the quadratic term is $\gamma_2 = 0.044$ ($t = 26.277$). The implied turning point lies within the post-event window, at $k^* = 3.584$, which means that market efficiency peaks approximately 10.5 months after the launch. This U-shaped dynamic (the inverse of the market efficiency) suggests that the net effect of LLMs on market efficiency is shaped by competing forces whose relative strength evolves as the ecosystem develops, which supports the model prediction 2.

5.3 Model Prediction 3: Performance Gap Across Investor Types

I now examine the distributional effect, namely whether LLMs have narrowed the performance gap between institutional and retail investors. A priori, the net effect is unclear, giving rise to two competing hypotheses.

On one hand, the performance gap could widen. Institutional investors who equipped with superior resources such as access to premium model tiers, bespoke fine-tuning, and advanced prompt-engineering may be better positioned to extract disproportionate value from the new technology, thereby amplifying their existing advantages (Blankespoor et al., 2024).

On the other hand, the gap may narrow through two powerful mechanisms. First, and

most directly, LLMs democratize access to sophisticated analytical capabilities. Retail investors can now process complex public information at a speed and scale that were previously the exclusive domain of institutions. This directly erodes the traditional institutional edge in information processing. Second, as the widespread use of LLMs makes the market more efficient overall, the pool of easily exploitable mispricings shrinks. This reduction in available alpha disproportionately impacts sophisticated institutions whose strategies are often geared toward identifying such inefficiencies. By contrast, retail investors may continue to benefit from access to private information. Such information can arise naturally: for example, individuals may obtain value-relevant insights through their employment in the same industry or with a customer or supplier (Boehmer et al., 2021). Together, these forces suggest that LLMs may level the informational playing field and compress the performance gap between institutional and retail traders.

To investigate this issue, I identify retail orders using the price-improvement algorithm of Boehmer et al. (2021), which provides a more comprehensive proxy for retail trading than data limited to a single broker, wholesaler, or exchange. Following Barber et al. (2024), I construct a standardized order-imbalance measure:

$$SOI_{id} = \frac{bp_{id} - bp_{md}}{\sigma(bp_{id})},$$

where bp_{id} is the buy proportion of stock i on day d (based on the number of trades), bp_{md} is the market-wide buy proportion, and $\sigma(bp_{id})$ is the standard deviation of the probability of a buy under the null, scaled by trade size v_{id}:

$$\sigma(bp_{id}) = \frac{\sqrt{v_{id}(1 - bp_{md})(bp_{md})}}{v_{id}}.$$

Intuitively, retail sentiment is inferred to be bullish when a stock's buy proportion exceeds the market-wide benchmark, with the precision of this inference adjusted for expected sampling variability.

Each day, stocks are sorted into five quantiles based on SOI_{id}. A long–short portfolio is then formed by taking a long position in the highest quantile (Q5) and a short position in the lowest quantile (Q1). Portfolios are constructed under both equal-weighted (EW) and trading-volume-weighted (TVW) schemes, and are rebalanced daily. The TVW variant in particular more closely mimics the behavior of retail traders, whose influence

scales with trading activity (Barber et al., 2024). An analogous procedure is applied to construct long–short portfolios based on institutional trades, identified following Lee and Radhakrishna (2000).

Table 8 reports the results. Panels A.1 and B.1 present unconditional estimates using the full cross-section. Retail investors' profits increase after the launch of LLMs, while institutional investors' performance declines, which supports the model prediction 3. The effect is most pronounced among high-exposure stocks, suggesting that LLM adoption disproportionately benefits retail traders. Figure 6 illustrates this pattern: the profitability improvement for retail investors between the pre- and post-ChatGPT periods is concentrated in high-exposure firms and is particularly evident in trading-volume-weighted portfolios. By contrast, in the low-exposure group, profitability shows little change.

To further highlight the shifting relative performance of retail versus institutional investors, I construct a spread portfolio that goes long the retail order-imbalance strategy (Q5–Q1 based on SOI_{retail}) and short the institutional order-imbalance strategy (Q5–Q1 based on $SOI_{inst20k}$), rebalanced daily. Figure 7 plots the cumulative profit from investing \$1 at the beginning of the sample. A clear upward trend emerges in the LLM era, suggesting that retail investors have gained a relative advantage over institutions.

This evidence should be interpreted with caution for at least three reasons. First, the retail identification captures only marketable retail orders, excluding retail limit orders that may carry opposite information and partially offset the measured effects (Boehmer et al., 2021). Second, the institutional proxy based on large trades (Lee and Radhakrishna, 2000) is subject to a known selection bias. Sophisticated institutions, such as hedge funds, increasingly rely on algorithmic strategies and order-splitting techniques to disguise their activity and minimize price impact (Campbell et al., 2009; Cready et al., 2014). As a result, this proxy may fail to capture the trading of the most informed institutions and instead disproportionately reflect the behavior of less sophisticated large traders. Third, the daily rebalancing of strategy portfolios implies potentially non-trivial transaction costs, making the reported performance best viewed as a stylized measure rather than an implementable trading strategy.

6 Discussion

6.1 Does the Variance Ratio Capture Mispricing?

The model and the empirical design use different efficiency objects because they operate under different observability assumptions. In the model, efficiency is defined as the distance between price and fundamental value, since the econometrician observes the fundamental V. In the data, however, fundamental value is unobserved, which is a general challenge in empirical research on market efficiency. Rösch et al. (2017) compare several commonly used measures, such as intraday return predictability (Chordia et al., 2005), variance ratios (Bessembinder, 2003), and Hasbrouck pricing errors (Hasbrouck, 1993), and show that they comove. They characterize the variance ratio as a measure of how closely prices adhere to a random-walk benchmark, making it a reduced-form proxy for the efficiency of short-horizon price discovery rather than a direct measure of the price-fundamental gap. This distinction is important. In a hypothetical market with no informed trading and little incorporation of new information, the variance ratio could mechanically appear low simply because prices move very little. But such a market would not be efficient in the informational sense relevant here; it would instead be uninformative. Accordingly, the interpretation in this paper is narrower: lower variance ratios are taken as evidence of faster and less distorted high-frequency price discovery, not as a literal estimate of the distance between price and fundamental value. This is the standard empirical compromise in settings where fundamentals are unobserved.

6.2 Post-Earnings Announcement Drift

A well-known alternative measure of market efficiency is post-earnings announcement drift (PEAD), the tendency for stock returns to continue drifting in the direction of an earnings surprise after the announcement. In an efficient market, earnings news should be incorporated into prices promptly, leaving little scope for such predictable post-announcement returns. If LLMs accelerate information processing, PEAD should attenuate after the introduction of ChatGPT. I therefore use PEAD as a robustness check on the main market-efficiency results, particularly because the variance ratio is a reduced-form proxy rather than a direct measure of mispricing. This test also speaks to

the mechanism: if LLMs improve efficiency by helping investors digest textual information more quickly, then delayed adjustment to earnings news should decline.

To test this prediction, I construct a standard long-short trading strategy based on standardized unexpected earnings (SUE). On each earnings announcement date, firms are sorted into quantiles by SUE, measured relative to the median analyst forecast from the I/B/E/S database. A daily rebalanced long-short portfolio is then formed by taking a long position in the highest-SUE quantile (Q5) and a short position in the lowest-SUE quantile (Q1). The performance of this strategy is evaluated by estimating its alpha relative to the Fama-French five-factor model augmented with momentum (Fama and French, 2015).

To avoid immediate announcement-day reactions and microstructure noise, positions are initiated two trading days after the announcement. Two post-announcement holding windows are considered: $[2, 30]$ and $[2, 60]$ trading days. When firms have multiple eligible announcements, positions naturally overlap, and portfolio weights are rebalanced daily using prevailing constituents. Each portfolio is constructed on both an equal-weighted (EW) and value-weighted (VW) basis, with the constraint that each rebalance involves at least a one-day holding period.

Table 9 presents the results. Unconditionally (Panels A.1 and B.1), the strategy's alpha declines across all specifications in the post-ChatGPT period. To link this decline to LLM adoption, I partition firms into high- and low-exposure groups, defined relative to the cross-sectional median exposure. The reduction in strategy profitability is concentrated almost entirely within the high-exposure group: for these firms, strategy alphas fall significantly across all holding periods and weighting schemes. By contrast, no consistent decline in alpha is observed for the low-exposure group. This pattern provides strong evidence that the attenuation of PEAD is driven by the accelerated incorporation of earnings news among firms more exposed to LLM analysis.[11] Figure 6 corroborates these findings visually, showing that the decline in strategy profitability between the pre- and post-ChatGPT periods is stark for high-exposure firms (particularly in value-weighted portfolios) but absent for their low-exposure counterparts.

[11] Although alphas decline in the high-exposure group after the LLM launch, the differences are not statistically significant in all specifications. The absence of a comparable pattern in the low-exposure group is nevertheless consistent with the interpretation that efficiency improvements are concentrated in firms with higher ChatGPT exposure.

The decline in the profitability of the SUE long–short strategy may arise from two distinct mechanisms associated with LLM adoption: (i) improved price discovery or (ii) overreaction due to crowd trading. To distinguish between these channels, I estimate the following event-time Difference-in-Differences regression:

$$CAR_{i,\tau,w} = \beta_1 SUE_i + \beta_2 (SUE_i \times Post_\tau) + \mathbf{X}_{i,\tau}\delta + \alpha_i + \lambda_\tau + \varepsilon_{i,\tau}, \tag{6}$$

where τ indexes the earning announcement event, and $CAR_{i,\tau,w}$ denotes the cumulative abnormal return (in excess of the market return) for firm i over window w. Firm fixed effects α_i and event-time fixed effects λ_τ are included, with $\mathbf{X}_{i,\tau}$ denoting a set of controls.

Table 10 reports the estimates for two event windows: the initial announcement window [0, 1] and the post-announcement drift window [2, 60]. The baseline coefficient β_1 is significantly positive in both specifications. As expected, this reflects the normal day-1 market reaction to earnings news in Column (1) and the existence of a significant PEAD in the pre-ChatGPT period in Column (2).

The key coefficient for distinguishing the two channels is the interaction term $SUE_i \times Post$. The overreaction hypothesis predicts an amplified initial price response to earnings news in the LLM era, which would imply a positive coefficient for the interaction term ($\beta_2 > 0$) in the announcement window. In contrast, the estimate in Column (1) is negative. While the immediate reaction to earnings remains intact (The sum of coefficients ($\beta_1 + \beta_2$) in Column (1) remains positive), the negative interaction term is inconsistent with the overreaction hypothesis. Instead, it suggests a more nuanced initial absorption of information. Crucially, the results for the drift window in Column (2) show a large, negative, and significant coefficient on the interaction term. The sum ($\beta_1 + \beta_2$) is no longer positive, consistent with the disappearance of PEAD in the post-ChatGPT period.

Overall, the evidence suggests that the reduced profitability of PEAD arises primarily from improved price discovery facilitated by LLMs, rather than from crowd trading or overreaction.

7 Conclusion

Generative AI is reshaping how financial markets process information. This paper develops a model in which LLM adoption increases the informativeness of order flow but also raises the correlation of trading errors, and tests the model's predictions using evidence from ChatGPT outages and its public launch. The results show that LLMs improve the speed of price discovery and narrow the performance gap between institutional and retail investors.

These findings also raise important questions for regulators and market designers. While LLMs can improve market quality, they may also create systemic risks when many participants rely on similar models and generate correlated trading patterns. An important challenge, therefore, is to design policies and market structures that preserve the informational benefits of LLMs while limiting the fragility that may arise from their widespread and concentrated use.

References

Alves, J. (2025). 'chatgpt, what stocks should i buy?' ai fuels boom in robo-advisory market. https://www.reuters.com/business/finance/chatgpt-what-stocks-should-i-buy-ai-fuels-boom-robo-advisory-market-2025-09-25/.

Barber, B. M., S. Lin, and T. Odean (2024). Resolving a paradox: Retail trades positively predict returns but are not profitable. *Journal of Financial and Quantitative Analysis 59*(6), 2547–2581.

Beckmann, L., H. Beckmeyer, I. Filippou, S. Menze, and G. Zhou (2024). Unusual financial communication: Chatgpt, earnings calls, and financial markets. *Olin Business School Center for Finance & Accounting Research Paper* (2024/02).

Bertomeu, J., Y. Lin, Y. Liu, and Z. Ni (2025). The impact of generative ai on information processing: Evidence from the ban of chatgpt in italy. *Journal of Accounting and Economics*, 101782.

Bessembinder, H. (2003). Trade execution costs and market quality after decimalization. *Journal of Financial and Quantitative Analysis 38*(4), 747–777.

Blankespoor, E., J. Croom, and S. M. Grant (2024). Generative ai and investor processing of financial information. *Available at SSRN 5053905*.

Boehmer, E., C. M. Jones, X. Zhang, and X. Zhang (2021). Tracking retail investor activity. *The Journal of Finance 76*(5), 2249–2305.

Boehmer, E. and E. K. Kelley (2009). Institutional investors and the informational efficiency of prices. *The Review of Financial Studies 22*(9), 3563–3594.

Brogaard, J., T. Hendershott, and R. Riordan (2014). High-frequency trading and price discovery. *The Review of Financial Studies 27*(8), 2267–2306.

Campbell, J. Y., T. Ramadorai, and A. Schwartz (2009). Caught on tape: Institutional trading, stock returns, and earnings announcements. *Journal of financial economics 92*(1), 66–91.

Cao, S. S., G. A. Karolyi, W. W. Xiong, and H. Xu (2026). Biodiversity entrepreneurship. *Review of Finance 30*(1), 43–86.

Chang, A., X. Dong, X. Martin, and C. Zhou (2023). Ai democratization, return predictability, and trading inequality. *Available at SSRN 4543999*.

Chen, H., A. Didisheim, L. Somoza, and H. Tian (2025). A financial brain scan of the llm. *arXiv preprint arXiv:2508.21285*.

Chen, J., G. Tang, G. Zhou, and W. Zhu (2025). Chatgpt and deepseek: Can they predict the stock market and macroeconomy? *arXiv preprint arXiv:2502.10008*.

Chen, S., L. Peng, and D. Zhou (2024). Wisdom or whims? decoding investor trading strategies with large language models. *Zicklin School of Business, Baruch College Working Papers*.

Cheng, Q., P. Lin, and Y. Zhao (2025). Does generative ai facilitate investor trading? early evidence from chatgpt outages. *Journal of Accounting and Economics*, 101821.

Chordia, T., R. Roll, and A. Subrahmanyam (2005). Evidence on the speed of convergence to market efficiency. *Journal of financial economics 76*(2), 271–292.

Cready, W., A. Kumas, and M. Subasi (2014). Are trade size-based inferences about traders reliable? evidence from institutional earnings-related trading. *Journal of Accounting Research 52*(4), 877–909.

Dou, W. W., I. Goldstein, and Y. Ji (2025). Ai-powered trading, algorithmic collusion, and price efficiency. Technical report, National Bureau of Economic Research.

Ebert, L. (2023). The rise of ai assistants: Hedge fund managers and chatgpt. https://globalmarkets.cib.bnpparibas/the-rise-of-ai-assistants-hedge-fund-managersand-chatgpt/.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The journal of Finance 25*(2), 383–417.

Fama, E. F. and K. R. French (2015). A five-factor asset pricing model. *Journal of financial economics 116*(1), 1–22.

Goldstein, I., Y. Xiong, and L. Yang (2025). Information sharing in financial markets. *Journal of Financial Economics 163*, 103967.

Grossman, S. J. and J. E. Stiglitz (1980). On the impossibility of informationally efficient markets. *The American economic review 70*(3), 393–408.

Guo, Y., S. Tong, T. Chen, S. Kumar, and S. Ouyang (2022). The impact of generative artificial intelligence on individual manual investment decisions: Empirical evidence from mutual funds. *Nanyang Business School Research Paper* (23-24).

Hansen, A. L. and S. Kazinnik (2024). Can chatgpt decipher fedspeak? *Available at SSRN 4399406*.

Hasbrouck, J. (1993). Assessing the quality of a security market: A new approach to transaction-cost measurement. *The Review of Financial Studies 6*(1), 191–212.

Hendershott, T., C. M. Jones, and A. J. Menkveld (2011). Does algorithmic trading improve liquidity? *The Journal of finance 66*(1), 1–33.

Hu, J., L. X. Liu, C. Y. Liu, H. Qu, and Y. Zhang (2025). Ceo turnover, sequential disclosure, and stock returns. *Review of Finance 29*(3), 887–921.

Huang, L., B. Qiao, and D. Zhou (2024). Are memes a sideshow: Evidence from wallstreetbets. *Available at SSRN 4785481*.

International Monetary Fund (2024). Global financial stability report. Technical Report October 2024.

Jha, M., J. Qian, M. Weber, and B. Yang (2024). Chatgpt and corporate policies. Technical report, National Bureau of Economic Research.

Ji, Z., N. Lee, R. Frieske, T. Yu, D. Su, Y. Xu, E. Ishii, Y. J. Bang, A. Madotto, and P. Fung (2023). Survey of hallucination in natural language generation. *ACM computing surveys 55*(12), 1–38.

Kacperczyk, M., J. Nosal, and S. Sundaresan (2025). Market power and price informativeness. *Review of Economic Studies 92*(3), 1955–1986.

Kang, H. and X.-Y. Liu (2023). Deficiency of large language models in finance: An empirical examination of hallucination. *arXiv preprint arXiv:2311.15548*.

Kim, A., M. Muhn, and V. Nikolaev (2023). Bloated disclosures: can chatgpt help investors process information? *arXiv preprint arXiv:2306.10224*.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, 1315–1335.

Lee, C. M. and B. Radhakrishna (2000). Inferring investor behavior: Evidence from torq data. *Journal of Financial Markets 3*(2), 83–111.

Li, K., F. Mai, R. Shen, C. Yang, and T. Zhang (2026). Dissecting corporate culture using generative ai. *The Review of financial studies 39*(1), 253–296.

Lo, A. W. and J. Ross (2024). Can chatgpt plan your retirement?: Generative ai and financial advice. *Harvard Data Science Review* (Special Issue 5).

Lopez-Lira, A. and Y. Tang (2023). Can chatgpt forecast stock price movements? return predictability and large language models. *arXiv preprint arXiv:2304.07619*.

Lv, L. (2024). The value of information from sell-side analysts. *arXiv preprint arXiv:2411.13813*.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica: journal of the Econometric Society*, 69–85.

Rösch, D. M., A. Subrahmanyam, and M. A. Van Dijk (2017). The dynamics of market efficiency. *The review of financial studies 30*(4), 1151–1187.

Serafeim, G. (2024). Climate solutions, transition risk, and stock returns.

Shaffer, M. and C. C. Wang (2024). Scaling core earnings measurement with large language models. *Available at SSRN*.

Sheng, J., Z. Sun, B. Yang, and A. L. Zhang (2024). Generative ai and asset management. *Available at SSRN 4786575*.

Shive, S. (2012). Local investors, price discovery, and market efficiency. *Journal of Financial Economics 104*(1), 145–161.

Stein, J. C. (2009). Presidential address: Sophisticated investors and market efficiency. *The Journal of Finance 64*(4), 1517–1548.

Xu, Y., Y. Xuan, and G. Zheng (2025). Ai agent misinformation when assisting financial decision-making: Early evidence from stock recommendations. *Available at SSRN 5651130*.

Zhang, Y., Y. Li, L. Cui, D. Cai, L. Liu, T. Fu, X. Huang, E. Zhao, Y. Zhang, Y. Chen, et al. (2025). siren's song in the ai ocean: A survey on hallucination in large language models. *Computational Linguistics*, 1–46.

Appendix A: Proofs

Proof of Proposition 1

Proof. Under perfect competition, the market maker's zero-profit condition implies

$$p(y) = \mathbb{E}[v \mid y] + \delta y.$$

Since

$$y = B(A)v + \varepsilon + A\eta,$$

the pair (v, y) is jointly Gaussian. Hence

$$\mathbb{E}[v \mid y] = \frac{\mathrm{Cov}(v, y)}{\mathrm{Var}(y)} y,$$

where

$$\mathrm{Cov}(v, y) = B(A)\sigma_v^2, \qquad \mathrm{Var}(y) = B(A)^2\sigma_v^2 + \sigma_\varepsilon^2 + A^2\sigma_\eta^2.$$

Substituting gives

$$\lambda(A) = \delta + \frac{B(A)\sigma_v^2}{B(A)^2\sigma_v^2 + \sigma_\varepsilon^2 + A^2\sigma_\eta^2}.$$

Given v, the informed trader solves

$$\max_x \ \mathbb{E}[(v - p)x \mid v].$$

Using $p = \lambda(x + u)$ and $\mathbb{E}[u \mid v] = A\phi v$, the objective becomes

$$\max_x \ (v - \lambda x - \lambda A\phi v)x.$$

The first-order condition is

$$v - 2\lambda x - \lambda A\phi v = 0,$$

which implies

$$x = \frac{1 - \lambda A\phi}{2\lambda} v.$$

Therefore,

$$\beta(A) = \frac{1 - \lambda(A)A\phi}{2\lambda(A)}.$$

This yields the stated equilibrium characterization. □

Proof of Proposition 2

Proof. Differentiate $G(B, A)$ with respect to A and B:

$$G_A(B, A) = \frac{\phi}{(2B - A\phi)^2} + \frac{2AB\sigma_v^2\sigma_\eta^2}{\left(B^2\sigma_v^2 + \sigma_\varepsilon^2 + A^2\sigma_\eta^2\right)^2},$$

and

$$G_B(B, A) = -\frac{2}{(2B - A\phi)^2} - \frac{\sigma_v^2\left(\sigma_\varepsilon^2 + A^2\sigma_\eta^2 - B^2\sigma_v^2\right)}{\left(B^2\sigma_v^2 + \sigma_\varepsilon^2 + A^2\sigma_\eta^2\right)^2}.$$

The expression for $G_A(B, A)$ is strictly positive. The condition in the proposition is equivalent to $G_B(B, A) < 0$. Since $G(B(A), A) = 0$, the implicit function theorem implies

$$B'(A) = -\frac{G_A(B(A), A)}{G_B(B(A), A)} > 0.$$

The condition $G_B(B(A), A) < 0$ also implies local uniqueness of the equilibrium branch. □

Proof of Proposition 3

Proof. Since

$$p = \mathbb{E}[v \mid y] + \delta y,$$

it follows that

$$v - p = \big(v - \mathbb{E}[v \mid y]\big) - \delta y.$$

Taking expectations after squaring both sides gives

$$M(A) = \operatorname{Var}(v \mid y) + \delta^2 \operatorname{Var}(y),$$

because the cross term vanishes by the orthogonality property of conditional expectation.

Since (v, y) are jointly Gaussian,

$$\operatorname{Var}(v \mid y) = \operatorname{Var}(v) - \frac{\operatorname{Cov}(v, y)^2}{\operatorname{Var}(y)}.$$

Using

$$\operatorname{Var}(v) = \sigma_v^2, \qquad \operatorname{Cov}(v, y) = B(A)\sigma_v^2, \qquad \operatorname{Var}(y) = S(A) + N(A),$$

we obtain

$$\operatorname{Var}(v \mid y) = \sigma_v^2 \frac{N(A)}{S(A) + N(A)}.$$

Substituting into $M(A) = \operatorname{Var}(v \mid y) + \delta^2 \operatorname{Var}(y)$ yields

$$M(A) = \sigma_v^2 \frac{N(A)}{S(A) + N(A)} + \delta^2\big(S(A) + N(A)\big).$$

Differentiating with respect to A gives

$$M'(A) = \sigma_v^2 \frac{N'(A)S(A) - N(A)S'(A)}{(S(A)+N(A))^2} + \delta^2\big(S'(A) + N'(A)\big).$$

At $A = 0$, $N(0) = \sigma_\varepsilon^2$ and $N'(0) = 0$, so

$$M'(0) = S'(0)\left[\delta^2 - \sigma_v^2 \frac{\sigma_\varepsilon^2}{(S(0)+\sigma_\varepsilon^2)^2}\right].$$

Since $B'(0) > 0$, we have $S'(0) > 0$. Hence $M'(0) < 0$ under the first inequality.

At $A = 1$, $N(1) = \sigma_\varepsilon^2 + \sigma_\eta^2$ and $N'(1) = 2\sigma_\eta^2$, so

$$M'(1) = \sigma_v^2 \frac{2\sigma_\eta^2 S(1) - (\sigma_\varepsilon^2 + \sigma_\eta^2)S'(1)}{(S(1)+\sigma_\varepsilon^2+\sigma_\eta^2)^2} + \delta^2\big(S'(1) + 2\sigma_\eta^2\big).$$

Thus $M'(1) > 0$ under the second inequality. By continuity of $M'(A)$, there exists some $A^* \in (0,1)$ such that $M'(A^*) = 0$. □

Proof of Proposition 4

Proof. Under the linear equilibrium,

$$x = \beta(A)v, \qquad p = \lambda(A)y, \qquad y = B(A)v + \varepsilon + A\eta,$$

where $B(A) \equiv \beta(A) + A\phi$. Substituting these expressions gives

$$\Pi_I(A) = \mathbb{E}\big[(v-p)\beta(A)v\big] = \beta(A)\,\mathbb{E}\big[v(v-p)\big].$$

Since

$$p = \lambda(A)\big(B(A)v + \varepsilon + A\eta\big),$$

it follows that

$$v - p = \big(1 - \lambda(A)B(A)\big)v - \lambda(A)\varepsilon - \lambda(A)A\eta.$$

Hence,

$$\mathbb{E}\big[v(v-p)\big] = \big(1-\lambda(A)B(A)\big)\mathbb{E}[v^2] - \lambda(A)\mathbb{E}[v\varepsilon] - \lambda(A)A\mathbb{E}[v\eta].$$

Because v, ε, and η are mutually independent,

$$\mathbb{E}[v\varepsilon] = 0, \qquad \mathbb{E}[v\eta] = 0, \qquad \mathbb{E}[v^2] = \sigma_v^2,$$

so

$$\Pi_I(A) = \beta(A)\big(1 - \lambda(A)B(A)\big)\sigma_v^2.$$

Using $B(A) = \beta(A) + A\phi$, this becomes

$$\Pi_I(A) = \beta(A)\big(1 - \lambda(A)\beta(A) - \lambda(A)A\phi\big)\sigma_v^2.$$

From the insider first order condition,

$$\beta(A) = \frac{1 - \lambda(A)A\phi}{2\lambda(A)},$$

which implies

$$\lambda(A)\beta(A) = \frac{1 - \lambda(A)A\phi}{2}.$$

Therefore,

$$1 - \lambda(A)\beta(A) - \lambda(A)A\phi = \frac{1 - \lambda(A)A\phi}{2} = \lambda(A)\beta(A).$$

Substituting this identity yields

$$\Pi_I(A) = \lambda(A)\beta(A)^2\sigma_v^2.$$

Finally, substituting

$$\beta(A) = \frac{1 - \lambda(A)A\phi}{2\lambda(A)}$$

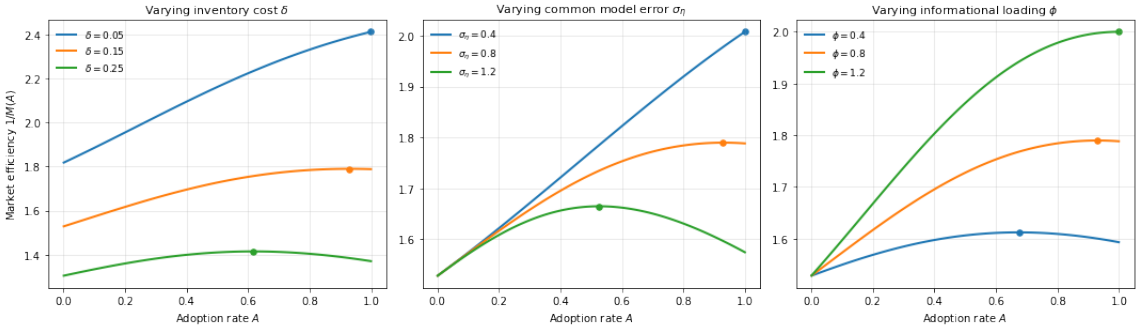
gives

$$\Pi_I(A) = \frac{\sigma_v^2}{4\lambda(A)}\big(1 - \lambda(A)A\phi\big)^2.$$

□

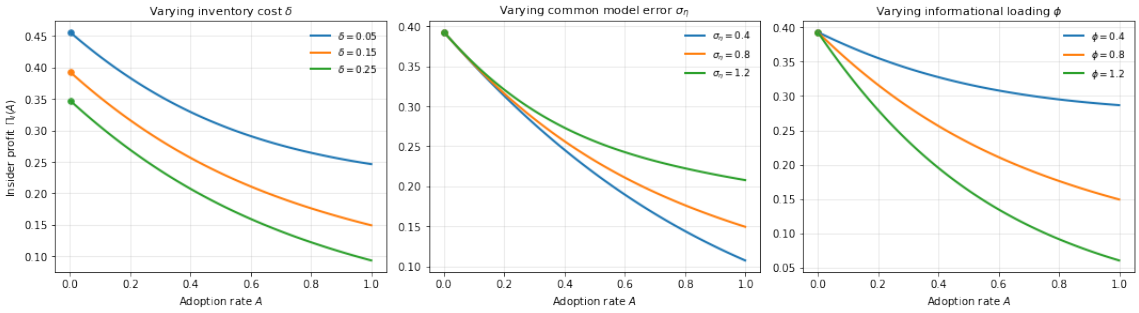
Appendix B: Figures and Tables

Figure 1: **Market Efficiency and LLM Adoption under Different Parameter Regions**



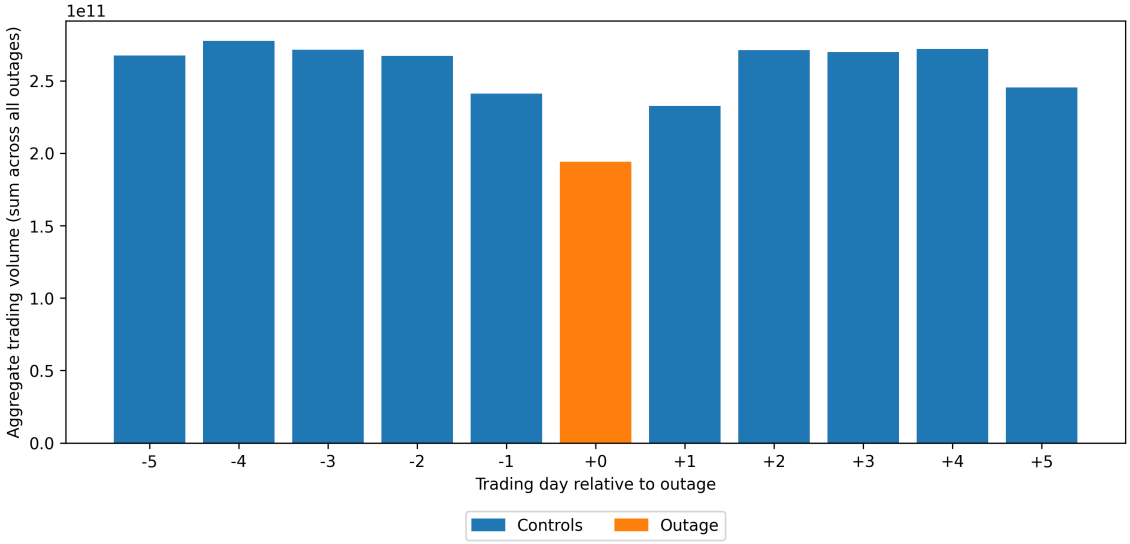
Notes: The figure plots market efficiency as a function of the adoption rate A under alternative values of the inventory cost δ, the common model error σ_η, and the informational loading ϕ, with other parameters held fixed at their baseline values. Baseline parameters are $\sigma_v = 1.0$, $\sigma_\varepsilon = 1.0$, $\sigma_\eta = 0.8$, $\phi = 0.8$, and $\delta = 0.15$.

Figure 2: **Insider profit and LLM Adoption under Different Parameter Regions**



Notes: The figure plots insider profit as a function of the adoption rate A under alternative values of the inventory cost δ, the common model error σ_η, and the informational loading ϕ, with other parameters held fixed at their baseline values. Baseline parameters are $\sigma_v = 1.0$, $\sigma_\varepsilon = 1.0$, $\sigma_\eta = 0.8$, $\phi = 0.8$, and $\delta = 0.15$.

Figure 3: **Trading Volume: Control vs. Outage Periods**

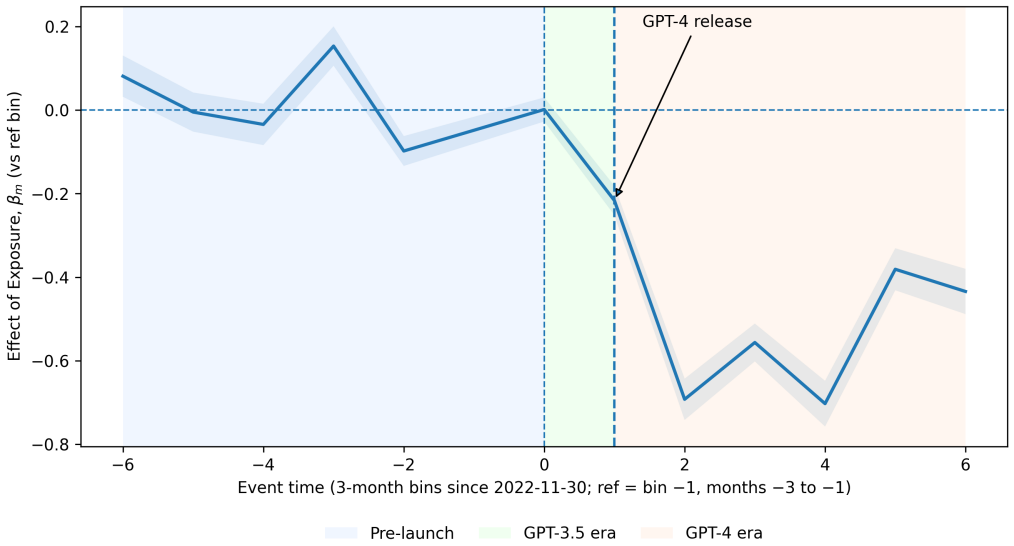


Notes: This figure plots the aggregate dollar trading volume of all stocks during outage periods at the 5-minute frequency, together with the corresponding intervals from the five trading days before and after each outage day. The sample covers November 30, 2022 to December 31, 2023 (results are unchanged when extending to 2024).

Figure 4: **Google Trend for LLM**

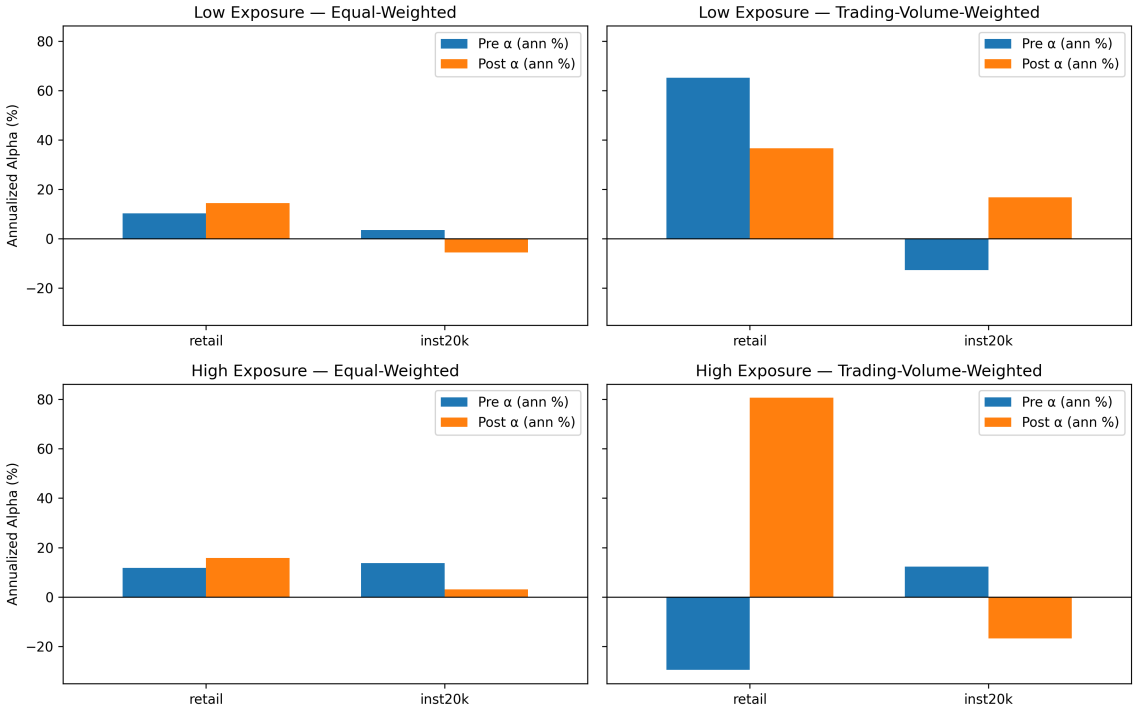
Notes: This figure plots the worldwide Google Trends index for the keyword *LLM* from January 2021 to December 2024.

Figure 5: **Dynamic Treatment Effects**



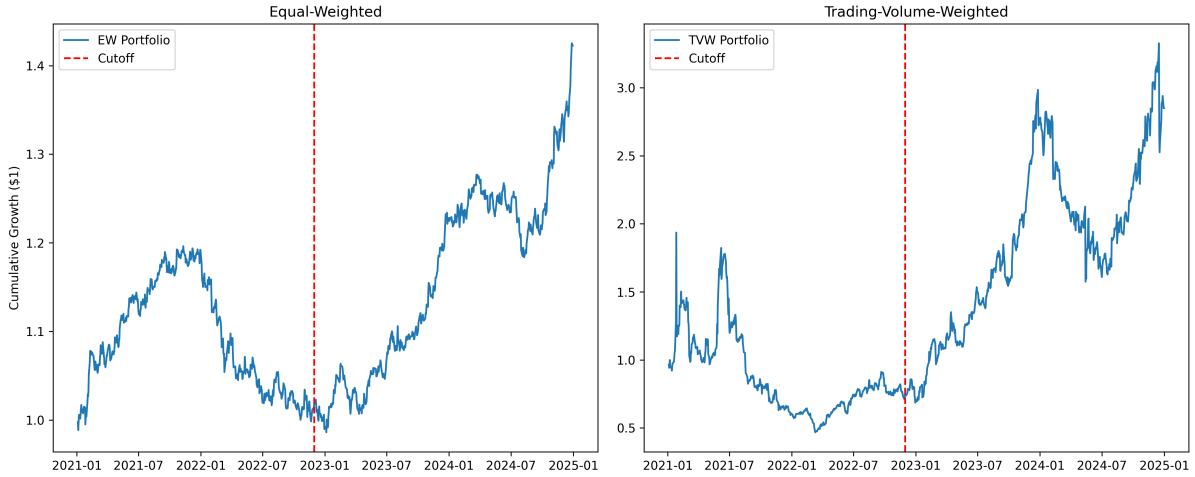
Notes: This figure plots the dynamic treatment effects by month (the coefficients β_k from Equation 4), measured relative to the ChatGPT launch. The leftmost point aggregates $k \leq -6$ (18+ months prior), and the rightmost point aggregates $k \geq 6$ (18+ months after). Negative β_k values indicate improved efficiency relative to the reference period, months $k = -3$ to -1. The shaded areas indicate different phases of ChatGPT adoption: the blue region denotes the pre-launch period, the green region marks the first three months after launch (the GPT-3.5 era), and the orange region represents the subsequent period (the GPT-4 era). The sample covers January 1, 2021 to December 31, 2023, with results unchanged when extended through 2024.

Figure 6: **Long-short Order Imbalance Portfolio Alphas by Groups**



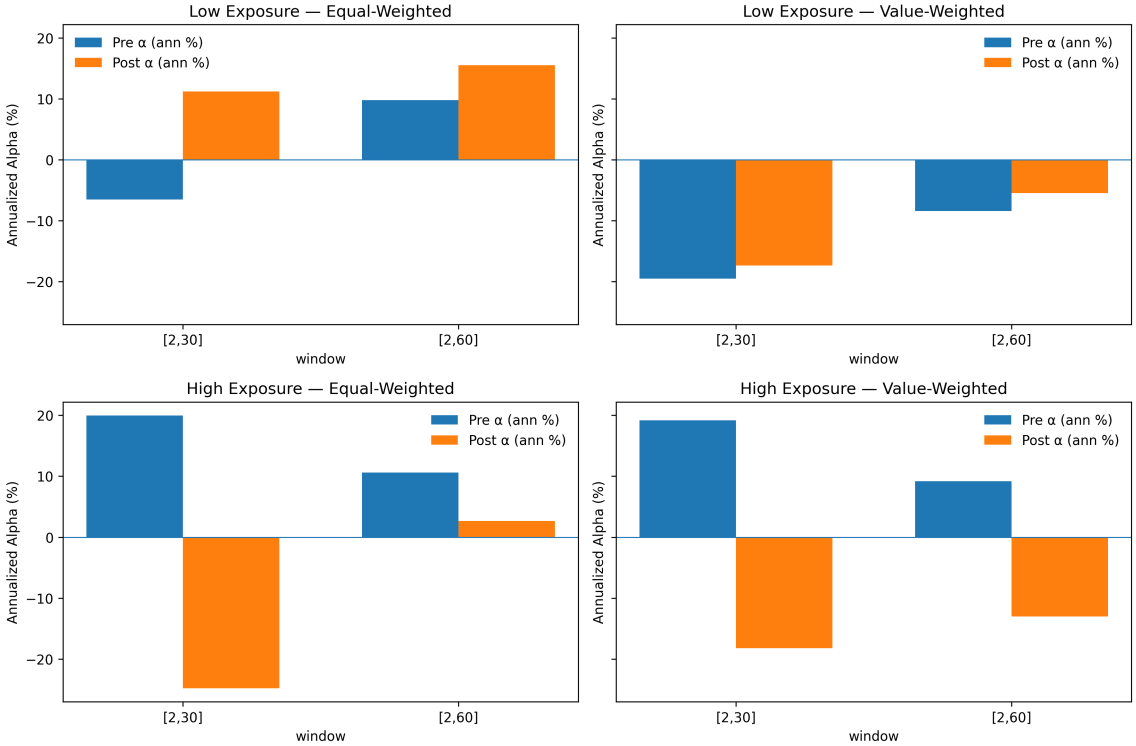
Notes: This figure plots the portfolio (annualized) alpha after adjusting the Fama French 5 Factors plus the Momentum Factor. "GPT Exposure" splits firms into Low vs High exposure (median split) and repeats the construction within each group. *exposure* is defined as the decile-based ranking of the decline in trading volume during outage events, averaged across all events, such that higher values indicate greater ChatGPT trading exposure. The whole sample period is from January 1, 2021 to December 31, 2024. Pre and Post refer to subsamples split at the break date (Nov 30, 2022).

Figure 7: **Cumulative Profits of Retail Minus Institutional Order-Imbalance Portfolios**



Notes: This figure plots the time series of cumulative returns from a long–short strategy that is long in the retail standardized order-imbalance portfolio (Q5–Q1) and short in the institutional standardized order-imbalance portfolio (Q5–Q1). The EW portfolio is equal-weighted, while the TVW portfolio is weighted by the number of trades. The initial investment is normalized to one unit.

Figure 8: **Long-short SUE Portfolio Alphas by Groups**



Notes: This figure plots the portfolio (annualized) alpha after adjusting the Fama French 5 Factors plus the Momentum Factor. "GPT Exposure" splits firms into Low vs High exposure (median split) and repeats the construction within each group. *exposure* is defined as the decile-based ranking of the decline in trading volume during outage events, averaged across all events, such that higher values indicate greater ChatGPT trading exposure. The whole sample period is from January 1, 2021 to December 31, 2024. Pre and Post refer to subsamples split at the break date (Nov 30, 2022).

Table 1: **ChatGPT Outage Events During the Market Hours**

date	start	end	name	duration (mins)
2023-02-14	14:30	14:33	Outage on ChatGPT	3.0
2023-02-27	10:00	10:40	ChatGPT outage	40.0
2023-03-15	14:14	14:20	ChatGPT Outage	6.0
2023-03-20	12:41	16:12	ChatGPT Web Interface Incident	211.0
2023-04-04	15:02	16:14	Increased errors on gpt-3-5-turbo, curie and da...	72.0
2023-06-30	14:41	15:20	Customer logins to ChatGPT are not available	39.0
2023-07-11	10:56	13:40	Elevated error rate for ChatGPT logins	164.0
2023-07-12	10:43	12:20	Elevated error rate for ChatGPT	97.0
2023-08-29	10:04	10:46	ChatGPT web & mobile UIs unavailable	42.0
2023-08-31	09:52	12:29	Degraded performance on ChatGPT	157.0
2023-09-12	09:19	10:34	Elevated error rates in ChatGPT	75.0
2023-09-13	09:04	09:58	Elevated error rates in ChatGPT	54.0
2023-09-15	15:56	16:43	High error rates for sampling requests to fine ...	47.0
2023-11-08	15:03	15:41	Periodic outages across ChatGPT and API	38.0
2023-11-23	09:31	09:37	Elevated ChatGPT Errors	6.0
2023-12-19	13:43	13:54	Elevated errors on GPT-V for ChatGPT models	11.0
2024-02-13	15:11	16:46	Elevated errors across multiple API endpoints a...	95.0
2024-02-14	10:03	10:08	Elevated error rate impacting ChatGPT (includin...	5.0
2024-02-13	10:03	11:30	Elevated errors on GPT-4V for ChatGPT and API	87.0
2024-02-14	13:20	13:38	Elevated error rate impacting ChatGPT (includin...	18.0
2024-02-28	09:47	11:29	Increased latency and errors affecting both Cha...	102.0
2024-03-13	15:29	15:56	Degraded performance in ChatGPT and GPT4 models	27.0
2024-04-11	10:05	10:42	Increased latency and error rate affecting gpt-...	37.0
2024-04-10	13:56	16:56	Elevated errors in ChatGPT	180.0
2024-05-09	13:21	14:02	Elevated errors on ChatGPT	41.0
2024-05-23	02:31	10:40	ChatGPT's ability to search the internet is bei...	489.0
2024-05-24	15:45	17:04	ChatGPT Not Loading for Some Users	79.0
2024-05-21	14:40	14:50	Partial outage for GPT-4o free users	10.0
2024-06-04	10:33	13:17	ChatGPT is unavailable for some users.	164.0
2024-07-17	11:58	13:29	Elevated error rates for ChatGPT	91.0
2024-08-05	14:53	16:15	Elevated errors in ChatGPT when using the brows...	82.0
2024-08-21	12:27	13:02	Elevated errors impacting logins to ChatGPT and...	35.0
2024-08-16	14:53	14:54	Elevated error rates for ChatGPT and Platform API	1.0
2024-08-28	11:19	11:44	Elevated error rates for ChatGPT	25.0
2024-08-28	14:16	16:36	Elevated errors rates in Assistants API and Cha...	140.0
2024-09-20	13:17	15:36	ChatGPT Is Not Loading for Some Users	139.0
2024-10-17	13:31	13:43	Users are currently unable to upload files in C...	12.0
2024-12-17	15:39	18:23	Degraded Performance in ChatGPT Advanced Voice ...	164.0
2024-12-26	14:00	17:05	High error rates for ChatGPT, APIs, and Sora	185.0

Notes: This table lists the Large Language Model (LLM) outage events identified for the sample period of November 30, 2022, to December 31, 2024. An event is included if it satisfies three criteria: (i) the event title references "GPT," (ii) the component status is officially designated as an "outage," and (iii) the outage period overlaps, at least partially, with U.S. market trading hours (9:30 AM to 4:00 PM ET). For each qualifying event, the table reports the start time, end time, total duration, and a description of the main issue.

Table 2: **Summary Statistics**

	count	mean	std	min	25%	50%	75%	max
Panel A: Market Inefficiency (5-min): Outage Group								
VR	962 244	0.529	0.336	0.008	0.257	0.505	0.750	1.726
VRorth	962 244	0.000	0.335	−0.648	−0.268	−0.022	0.214	1.212
Panel B: Market Inefficiency (5-min): Control Group I								
VR	8 565 268	0.527	0.337	0.008	0.255	0.502	0.750	1.722
VRorth	8 565 268	0.000	0.336	−0.644	−0.268	−0.022	0.214	1.240
Panel C: Market Inefficiency (5-min): Control Group II								
VR	4 357 720	0.526	0.329	0.008	0.259	0.506	0.750	1.661
VRorth	4 357 720	0.000	0.328	−0.646	−0.264	−0.019	0.215	1.150
Panel D: Market Inefficiency (Daily)								
VR	2 897 642	0.321	0.229	0.005	0.114	0.280	0.512	0.801
Panel E: Daily Control Variables								
espread	2 907 040	0.303	0.552	0.014	0.056	0.109	0.256	4.233
volatility	2 899 537	13.181	52.344	0.015	0.162	0.556	2.950	558.308
ret	2 908 420	0.000	0.029	−0.113	−0.013	0.000	0.013	0.130
prc	2 908 420	61.184	86.203	5.000	12.450	28.900	69.800	482.950
turnover	2 908 420	0.014	0.304	0.000	0.003	0.006	0.010	104.134
InstOwnership	2 797 810	0.711	0.262	0.000	0.563	0.788	0.908	1.091
InfoEvents	2 902 905	12.339	8.635	0.000	6.000	11.000	16.000	42.000
complexity	2 744 731	20.547	4.917	13.081	15.325	20.827	25.103	28.896
logsize	2 908 420	14.337	1.949	8.275	12.894	14.284	15.591	18.984
VIX	1281	19.908	5.386	12.256	15.850	18.840	22.990	35.214

Notes: This table reports descriptive statistics for the main variables. The outage group consists of 962,244 stock–date–5-minute interval observations from 3,162 stocks during November 30, 2022 to December 31, 2024. VR is the variance ratio; $VRorth$ is the variance ratio orthogonalized with respect to 5-minute trading volume and quoted spread. Control Group I consists of the same intervals in the five trading days before and after each outage day, while Control Group II consists of non-outage intervals within outage days. Panel D reports daily firm-level VR, based on all trading days from January 1, 2021 to December 31, 2024. Panel E reports daily firm-level control variables (except the VIX) on these trading days. *espread* is the average percentage effective spread across trades k, defined as $espread_k = 2D_k(P_k - M_k)/M_k$, where P_k is the trade price, M_k the bid–ask midpoint, and $D_k \in \{+1, -1\}$ indicates buyer- or seller-initiated trades. *volatility* is intraday trade-based volatility (scaled by 10^6); *ret* and *prc* are daily return and price; *turnover* is the ratio of dollar volume to market capitalization; $InstOwnership$ is institutional ownership in percent; $InfoEvents$ counts firm-specific information events in the past 90 days; *complexity* is the Flesch–Kincaid grade level of 10-K/10-Q filings; *logsize* is the natural logarithm of market capitalization (scaled by 1,000); and VIX is the CBOE Volatility Index. All variables are winsorized at the 1% tails.

Table 3: **ChatGPT Outage and Market Efficiency**

	VR I	VR II	VRorth I	VRorth II
outage	0.132***	0.232***	0.159***	0.294***
	(0.037)	(0.039)	(0.037)	(0.039)
VR_L1	0.019***	0.018***	0.015***	0.012***
	(0.001)	(0.002)	(0.001)	(0.002)
espread_L1	1.126***	1.360***	-3.080***	-3.815***
	(0.225)	(0.301)	(0.241)	(0.325)
volatility_L1	-0.005***	-0.002	0.007***	0.013***
	(0.002)	(0.002)	(0.002)	(0.003)
ret_L1	-0.075	-4.096***	-0.564	-4.513***
	(0.461)	(0.688)	(0.463)	(0.685)
prc_L1	-0.001	-0.002	-0.004**	-0.004**
	(0.002)	(0.002)	(0.002)	(0.002)
InstOwnership_L1	-5.072***	-6.157***	-4.689***	-6.006***
	(0.743)	(0.732)	(0.717)	(0.719)
InfoEvents_L1	0.001	-0.006	0.008*	-0.001
	(0.004)	(0.004)	(0.004)	(0.005)
VIX_L1	0.083***	0.017**	0.075***	-0.008
	(0.006)	(0.008)	(0.006)	(0.008)
complexity_L1	0.066***	0.178***	0.068***	0.195***
	(0.011)	(0.014)	(0.011)	(0.014)
turnover_L1	-0.012	0.061***	-0.032	0.096***
	(0.041)	(0.010)	(0.045)	(0.016)
logsize_L1	-0.115	-0.498***	0.100	-0.370*
	(0.176)	(0.193)	(0.171)	(0.191)
ret[-5,-2]	-0.225	1.708***	-0.316	1.808***
	(0.245)	(0.327)	(0.245)	(0.325)
Time interval FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	9527512	5319964	9527512	5319964
R2 (%)	0.023	0.033	0.020	0.032

Notes: This table reports the regression results from Equation 1. Columns (1) and (3) compare the outage group with Control Group I, while Columns (2) and (4) use Control Group II. The suffix $L1$ denotes a one-day lag. Variance ratios (VR, $VRorth$, and VR_L1) are scaled by 100. *espread* is the average percentage effective spread of all trades (scaled by 100). *volatility* measures intraday trade-based volatility (scaled by 10^6). *ret* and *prc* are daily return and price, respectively. *turnover* is the ratio of dollar trading volume to market capitalization. *InstOwnership* denotes institutional ownership as a percentage. $InfoEvents$ counts firm-specific information events in the preceding 90 days. *complexity* is measured by the Flesch–Kincaid grade level of the firm's 10-K or 10-Q filings. *logsize* is the natural logarithm of market capitalization (scaled by 1,000). *cumret_t5_t2* denotes cumulative returns from day $t-5$ to day $t-2$. VIX is the market-level Volatility Index. The sample covers the period November 30, 2022 to December 31, 2024. All specifications include 5 minute time-interval and firm fixed effects. All variables are winsorized at the 1% tails. Standard errors (in parentheses) are clustered at the firm level. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table 4: **Firm Heterogeneity in Variance Ratio Changes**

	dVRI	dVRII
$\overline{espread\_L1}$	-0.583***	-0.765***
	(0.056)	(0.067)
$\overline{volatility\_L1}$	0.005***	0.008***
	(0.001)	(0.001)
$\overline{ret\_L1}$	-0.020	0.983*
	(0.463)	(0.537)
$\overline{prc\_L1}$	-0.000	-0.000**
	(0.000)	(0.000)
$\overline{InstOwnership\_L1}$	0.047***	0.108***
	(0.017)	(0.022)
$\overline{InfoEvents\_L1}$	0.000	0.000
	(0.001)	(0.001)
$\overline{VIX\_L1}$	0.174***	0.229***
	(0.032)	(0.073)
$\overline{complexity\_L1}$	0.009**	0.011**
	(0.004)	(0.005)
$\overline{turnover\_L1}$	0.047*	0.064***
	(0.027)	(0.022)
$\overline{logsize\_L1}$	-0.001	0.025***
	(0.003)	(0.004)
$\overline{ret[-5,-2]}$	0.222	0.727***
	(0.209)	(0.248)
Within-Firm Deviations Controls	Yes	Yes
Event Time FE	Yes	Yes
N	958063	961297
R2 (%)	0.064	0.209

Notes: This table reports the regression results from Equation 2. Columns (1) use the outage group and Control Group I, while column (2) use the outage group and Control Group II. dVR is defined as the decide-based ranking (from 1 to 10) of the increase in variance ratio during the outage events, such that a higher rank denotes a larger increase in variance ratio during the outage. *espread* is the average percentage effective spread of all trades (scaled by 100); *volatility* measures intraday trade-based volatility (scaled by 10^6); *ret* and *prc* are daily return and price, respectively; *turnover* is the ratio of dollar trading volume to market capitalization; $InstOwnership$ denotes institutional ownership as a percentage; $InfoEvents$ counts firm-specific information events in the preceding 90 days; *complexity* is measured as the Flesch–Kincaid grade level of the firm's 10-K or 10-Q filings; *logsize* is the natural logarithm of market capitalization (scaled by 1,000); $ret[5,2]$ denotes the cumulative returns from day $t-5$ to day $t-2$. The suffix $L1$ denotes the one-day lag. The bar denotes the average across all events. The sample covers the period from November 30, 2022 to December 31, 2024. All specifications include event time firm fixed effects and the Within-Firm Deviations Controls which accounts for the within firm difference across different events. All the independent variables are winsorized at the 1% tails. Standard errors (in parentheses) are clustered at the firm level. ***, ** and * denote significant level 1%, 5% and 10% level respectively.

Table 5: **ChatGPT Outage and Trading Volume**

	TV	TVRetail	TVNonRetail
outage	-17625.499***	-87.351	-17538.148***
	(2227.634)	(209.878)	(2092.177)
ret[-5,-2]	-49104.949	-3678.894	-45426.056
	(46788.290)	(5185.023)	(42505.479)
espread_L1	18993.440*	4178.578***	14814.863
	(10547.943)	(1517.478)	(9150.246)
volatility_L1	-145.950***	-21.631***	-124.319***
	(55.068)	(7.800)	(48.135)
ret_L1	114685.971**	21138.946***	93547.026*
	(58461.066)	(7104.524)	(52521.346)
prc_L1	56.938	-6.073	63.010
	(574.848)	(56.490)	(523.339)
InstOwnership_L1	379515.867***	34225.881***	345289.986***
	(77344.618)	(10292.910)	(69323.033)
InfoEvents_L1	4043.558***	285.152***	3758.406***
	(707.701)	(68.374)	(654.254)
VIX_L1	8484.149***	376.224***	8107.925***
	(863.201)	(79.347)	(803.808)
complexity_L1	-3527.276***	-616.781***	-2910.494**
	(1227.905)	(125.393)	(1131.262)
turnover_L1	47350.101	6623.844	40726.258
	(41007.317)	(5820.752)	(35226.144)
logsize_L1	195488.058***	15546.914***	179941.143***
	(34663.849)	(3851.735)	(31309.552)
Time interval FE	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes
N	5112380	5112380	5112380
R2 (%)	0.405	0.238	0.395

Notes: This table reports the regression results from

$$TV_{itd} = \beta\, Outage_t + \mathbf{X_{i,d-1}}\delta + \alpha_i + \gamma_t + \varepsilon_{it}$$

The sample period runs from November 30, 2022 to December 31, 2023 (results are unchanged when including 2024). The sample consists of the outage group and observations from the corresponding intervals in the five trading days preceding and following each outage day. The suffix $L1$ denotes a one-day lag. Trading volume is measured as the dollar value of total trades in a 5-minute interval. Retail trading is identified using the algorithm of Boehmer et al. (2021), and non-retail trading volume is defined as the difference between total trading and retail trading. *espread* is the average percentage effective spread of all trades (scaled by 100); *volatility* measures intraday trade-based volatility (scaled by 10^6); *ret* and *prc* are daily return and price, respectively; *turnover* is the ratio of dollar trading volume to market capitalization; $InstOwnership$ denotes institutional ownership as a percentage; $InfoEvents$ counts firm-specific information events in the preceding 90 days; *complexity* is measured as the Flesch–Kincaid grade level of the firm's 10-K or 10-Q filings; *logsize* is the natural logarithm of market capitalization (scaled by 1,000); $ret[5,2]$ denotes the cumulative returns from day $t-5$ to day $t-2$; VIX is the Market-level Volatility Index. All specifications include time interval and firm fixed effects. All the variables are winsorized at the 1% tails. Standard errors (in parentheses) are clustered at the firm level. ***, ** and * denote significant level 1%, 5% and 10% level respectively.

Table 6: **Launch of ChatGPT and Market Efficiency**

	VR	VR (DiD)
Post	-0.812***	0.488***
	(0.078)	(0.189)
exposure		-0.576***
		(0.022)
Post X exposure		-0.164***
		(0.034)
VR_L1	0.290***	0.581***
	(0.003)	(0.001)
ret[-5,-2]	-1.225***	-0.369***
	(0.144)	(0.127)
espread_L1	2.518***	5.464***
	(0.152)	(0.041)
volatility_L1	-0.010***	-0.033***
	(0.001)	(0.000)
ret_L1	-0.923***	0.452*
	(0.260)	(0.269)
prc_L1	0.020***	0.024***
	(0.002)	(0.000)
InstOwnership_L1	-9.228***	-5.717***
	(0.676)	(0.037)
InfoEvents_L1	-0.049***	-0.053***
	(0.003)	(0.001)
VIX_L1	-0.181***	-0.109***
	(0.004)	(0.002)
complexity_L1	0.040***	0.032***
	(0.008)	(0.003)
turnover_L1	-0.958***	-1.209***
	(0.164)	(0.214)
logsize_L1	-2.149***	-2.982***
	(0.164)	(0.008)
N	2487950	2487950
R2	0.112	0.732

Notes: This table reports the regression results from Equation 3. Columns (1) does not include the interaction and exposure terms. *Exposure* is defined as the decile-based ranking of the decline in trading volume during outage events, averaged across all events, such that higher values indicate greater ChatGPT trading exposure. The suffix $L1$ denotes the one-day lag. Variance ratios are scaled by 100; *espread* is the average percentage effective spread of all trades (scaled by 100); *volatility* measures intraday trade-based volatility (scaled by 10^6); *ret* and *prc* are daily return and price, respectively; *turnover* is the ratio of dollar trading volume to market capitalization; $InstOwnership$ denotes institutional ownership as a percentage; $InfoEvents$ counts firm-specific information events in the preceding 90 days; *complexity* is measured as the Flesch–Kincaid grade level of the firm's 10-K or 10-Q filings; *logsize* is the natural logarithm of market capitalization (scaled by 1,000); $ret[-5,-2]$ denotes the cumulative returns from day $t-5$ to day $t-2$; VIX is the Market-level Volatility Index. The sample covers the period from Jan 1, 2021 to December 31, 2024. All the variables are winsorized at the 1% tails. ***, ** and * denote significant level 1%, 5% and 10% level respectively.

Table 7: **Test of U-shaped Dynamics**

	Dependent variable: VR_{id}	
	Coefficient	t-statistic
$k_{\text{post}} \times Exposure_i$	-0.315***	-26.142
$k^2_{\text{post}} \times Exposure_i$	0.044***	26.277
Turning point k^*	3.584	

Notes: This table reports the coefficients from the post-event polynomial in event time based on regression 5. The implied turning point is $k^* = -\gamma_1/(2\gamma_2)$. The sample covers the period from January 1, 2021 to December 31, 2024. All variables are winsorized at the 1% tails. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. The specification includes month and firm fixed effects.

Table 8: **Portfolio Alphas from Order Imbalance Long-Short Strategy**

Panel A: Equal-Weighted Portfolios

A1. Unconditional

Type	Pre α	Post α	$\Delta\alpha$
Retail	11.37%**	15.50%***	4.13%
Inst20k	9.84%***	-0.63%	-10.47%**

A2. By GPT Exposure (Low vs High)

	Low Exposure			**High Exposure**		
Type	Pre α	Post α	$\Delta\alpha$	Pre α	Post α	$\Delta\alpha$
Retail	10.32%**	14.46%**	4.14%	11.71%**	15.77%***	4.05%
Inst20k	3.50%	-5.50%	-9.00%	13.77%***	3.11%	-10.66%**

Panel B: Trading-Volume-Weighted Portfolios

B1. Unconditional

Type	Pre α	Post α	$\Delta\alpha$
Retail	13.92%	75.36%**	61.44%
Inst20k	3.88%	0.12%	-3.76%

B2. By GPT Exposure (Low vs High)

	Low Exposure			**High Exposure**		
Leg	Pre α	Post α	$\Delta\alpha$	Pre α	Post α	$\Delta\alpha$
Retail	65.21%	36.56%	-28.65%	-29.53%	80.67%*	110.20%*
Inst20k	-12.66%	16.81%	29.47%	12.26%	-16.72%	-28.98%

Notes: This table reports annualized alphas (in %) from FF6 regressions of long-short portfolios formed on standard order imbalance (SOI) quantiles, split by investor type: retail (Boehmer et al., 2021) vs. institutional (Lee and Radhakrishna, 2000)), and GPT trading exposure. "Unconditional" uses the full cross-section; "By GPT Exposure" divides firms into Low vs High exposure groups and re-estimates the portfolios within each group. *exposure* is defined as the decile-based ranking of the decline in trading volume during outage events, averaged across all events, such that higher values indicate greater ChatGPT trading exposure. Trading-Volume-Weighted Portfolios weight each stock in the quantiles by the number of trades. The whole sample period is from January 1, 2021 to December 31, 2024. Pre and Post refer to subsamples split at the break date (Nov 30, 2022). $\Delta\alpha$ is the difference between post- and pre-period alphas. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Table 9: **Portfolio Alphas from SUE Long-Short Strategy**

Panel A: Equal-Weighted Portfolios

A1. Unconditional

Window	Overall α	Pre α	Post α	$\Delta\alpha$
$[2, 30]$	2.85%	20.75%**	-10.64%	-26.02%*
$[2, 60]$	9.49%	13.12%**	6.27%	-6.06%

A2. By GPT Exposure (Low vs High)

	Low Exposure				**High Exposure**			
Window	Overall α	Pre α	Post α	$\Delta\alpha$	Overall α	Pre α	Post α	$\Delta\alpha$
$[2, 30]$	2.21%	-6.50%	11.23%	18.96%	-5.63%	19.97%	-24.79%	-37.33%*
$[2, 60]$	12.98%**	9.79%	15.55%	5.25%	6.48%	10.63%	2.66%	-7.21%

Panel B: Value-Weighted Portfolios

B1. Unconditional

Window	Overall α	Pre α	Post α	$\Delta\alpha$
$[2, 30]$	-11.79%	2.15%	-21.52%**	-23.17%
$[2, 60]$	0.78%	3.02%	-0.95%	-3.86%

B2. By GPT Exposure (Low vs High)

	Low Exposure				**High Exposure**			
Window	Overall α	Pre α	Post α	$\Delta\alpha$	Overall α	Pre α	Post α	$\Delta\alpha$
$[2, 30]$	-18.66%*	-19.51%	-17.38%	2.64%	-1.75%	19.17%	-18.23%	-31.40%
$[2, 60]$	-5.56%	-8.41%	-5.49%	3.19%	-3.30%	9.16%	-13.03%	-20.33%

Notes: This table reports annualized alphas from FF6 factors (Fama-French five factors plus a momentum factor) regressions of long-short portfolios formed on SUE quantiles. Windows are in trading days after the earnings announcement. "Unconditional" uses the full cross-section; "By GPT Exposure" splits firms into Low vs High exposure (median split) and repeats the construction within each group. *exposure* is defined as the decile-based ranking of the decline in trading volume during outage events, averaged across all events, such that higher values indicate greater ChatGPT trading exposure. The whole sample period is from January 1, 2021 to December 31, 2024. Pre and Post refer to subsamples split at the break date (Nov 30, 2022). $\Delta\alpha$ is the difference between post- and pre-period alphas. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 10: **Event-time Differnece-in-differences Regression**

	CAR[1]	CAR[2,60]
SUE	0.337***	0.173**
	(0.075)	(0.083)
SUE×Post	-0.255***	-0.211**
	(0.084)	(0.094)
espread	-0.002	-0.015*
	(0.005)	(0.009)
volatility	-0.000	0.000
	(0.000)	(0.000)
ret	0.021**	0.039
	(0.011)	(0.025)
prc	0.000	0.000***
	(0.000)	(0.000)
InstOwnership	-0.039***	-0.130***
	(0.012)	(0.030)
InfoEvents	0.000	0.000
	(0.000)	(0.000)
complexity	-0.000	0.002***
	(0.000)	(0.001)
turnover	0.025	0.029
	(0.016)	(0.025)
logsize	-0.033***	-0.123***
	(0.003)	(0.007)
Event Time FE	Yes	Yes
Firm FE	Yes	Yes
N	28642	30284
R2	0.025	0.057

Notes: This table reports the regression results from Equation 6. The dependent variable in Column (1) is the abnormal return relative to the market on the first trading day following the earnings announcement. The dependent variable in Column (2) is the cumulative abnormal return relative to the market from day 2 to day 60 after the announcement. The variable SUE denotes standardized unexpected earnings, measured relative to the median analyst forecast. All control variables are winsorized at the 1% tails: effective spread (*espread*, scaled by 100), volatility (scaled by 10^6), returns (*ret*), price (*prc*), turnover, institutional ownership percentage ($InstOwnership$), number of firm-specific events in the past 90 days ($InfoEvents$), report complexity, firm size (*logsize*, scaled by 1,000). The sample covers January 1, 2021 to December 31, 2024. All specifications include event time and firm fixed effects. Standard errors (in parentheses) are clustered at the firm level. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.